THREE ESSAYS ON THE APPALACHIAN REGION

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy

in the Graduate School of The Ohio State University

By

Robert William Baumann, M.A.

* * * * *

The Ohio State University

2003

Dissertation Committee

Professor Patricia Reagan, Advisor

Professor Stephen Cosslett

Professor Joseph Kaboski

Professor Bruce Weinberg

Approved by

____

Advisor

Department of Economics

ABSTRACT

The literature has largely ignored white poverty, perhaps because discrimination, particularly institutionalized discrimination, is not a factor driving the process. White poverty is also more easily ignored because it is rarely concentrated to the same extent as central city black poverty. The major exception to the diffusion of white poverty is Appalachia, a region that for decades has experienced the greatest concentration of white poverty in the U.S. My goal is to evaluate the role played by differences in human capital and economic opportunity on the outcomes of both Appalachian and non-Appalachian poor whites. I argue that a main determinant in the Appalachia's relative economic deprivation is lower levels of overall human capital and economic opportunity than the rest of the U.S.

The above analysis will be divided into three essays. In the first essay, I estimate a three-stage wage equation model with two additional endogenous regressors, migration and employment, to determine how much of the Appalachian wage gap can be explained by the effects of human capital and local conditions and to estimate returns to migration using data from the National Longitudinal Surveys of Youth, 1979. I find differences in human capital and economic opportunity account for all of the differences in employment and most of the difference in wages. Migration offers small absolute returns for Appalachians, but does not raise wages to the level of poor white non-Appalachians. The second essay uses the same data and model but simultaneously estimates the parameters using Maximum Simulated Likelihood (MSL) to achieve higher efficiency than the multi-step method. Human capital differences again explain almost all of the wage and employment gaps. Migration offers no return using MSL in this setting. The final essay decomposes the wage gap between Appalachia and the rest of the country into quantities and prices of human capital and industry and occupation shares, Data are taken from the Integrated Micro Public Use (IPUMS) census data project. The large increases in the wage gap during the 1980s were largely caused by changes in income inequality and skill prices unfavorable to Appalachians.

ACKNOWLEDGMENTS

I wish to express my sincere gratitude to my advisor Patricia Reagan, whose guidance, support, and patience throughout this process made this dissertation possible. Thank you for going far above and beyond any expectations. You have made a profound mark on my life. Bruce Weinberg, Joseph Kaboski, and Stephen Cosslett also deserve recognition for their efforts as members of my dissertation committee. Each offered valuable insight that greatly improved the finished product. I am also grateful to Randy Olsen, Raphael Solomon, and the participants of the Labor Lunch and Microeconomics Seminar at Ohio State for their helpful comments. I am also grateful to the Center for Human Resource Research, particularly Randy Olsen and Rosella Gardecki, for providing an excellent working environment over the past four years. Thank you for your support, flexibility, and providing valuable training.

I am also fortunate to have encouragement from many friends and family. Special thanks are given to Christina Hochstettler, Allyson Hanson, Kristy Burrell, Garret Burrell, Danielle Levulis, Brandon Levulis, David Burrell, Logan Burrell, Matthew King, Dan Landis, Drew Weininger, Anthony Buehrer, Michael Dal Pra, Jefferson W. Nelson, Tom Stover, W. Skenyon Hansen, Jeremy Seidt, and Cara Orr. Thank you for your invaluable moral support.

Above all, I am grateful to my parents, Dan and Bobbie Baumann, who have provided endless support and a wealth of opportunity. This work is a testament to your tireless efforts and dedication over the years. Thank you for continually putting me in position to succeed.

VITA

February 4, 1976	Born in Columbus, OH
1998	B.A. Economics, Mathematics, Bluffton College, Bluffton, OH
2000	M.A., Economics, The Ohio State University
1998, Summer 2001, Summer 2002	Graduate Teaching Associate, Department of Economics, The Ohio State University
Fall 2002	Dice Fellowship
Winter 2003	PEGS Fellowship
1999-present	Graduate Research Associate, Center for Human Resource Research, The Ohio State University

FIELDS OF STUDY

Major Field: Economics

TABLE OF CONTENTS

Abstract ii

Acknowledgments iv

Vita vi

List of Tables ix

1. Introduction 1

    1.2 Overview of Economic Conditions in Appalachia 8

2. A Comparison of Employment, Earnings, and Migration of Economically Disadvantaged Whites in Appalachia and the Rest of the United States 13

    2.2 Model 14

    2.3 Data 21

    2.4 Results 26

    2.5 Conclusions 32

3. A Comparison of Employment, Earnings, and Migration of Economically Disadvantaged Whites Using Maximum Simulated Likelihood 47

    3.2 MSL method 49

        3.2.1 The Migration Decision 51

        3.2.2 The Employment Decision 52

3.3.4 Identification 55
3.4 Results 56
3.5 Conclusions 59
4. The Appalachian Wage Gap, 1940-1990 66
4.2 Data 71
4.3 Decomposition Methodology 78
4.4 Decomposition Results 84
4.5 Demand Shifts 87
4.6 Conclusions 89
List of References 105
Endnotes 109

LIST OF TABLES

Table 1.1 - Poverty, Unemployment, Income and Education Statistics 12

Table 2.1 - Sample Selection Criteria 34

Table 2.2 - Variable Means Used in Migration Equation 35

Table 2.3 - Variable Means Used in Employment Equation 37

Table 2.4 - Variable Means Used in Wage Equation... 39

Table 2.5 - Migration Probit 41

Table 2.6 - Employment Probit... 43

Table 2.7 - Log Wage Equations 45

Table 3.1 - MSL Estimates, Wage Equation... 62

Table 3.2 - MSL Estimates, Employment Equation 63

Table 3.3 - MSL Estimates, Migration Equation.. 64

Table 3.4 - MSL Estimates, Random effects structure 65

Table 4.1 - U.S. Mean Real Log Hourly Wages Levels by Year 92

Table 4.2 – Means of Observable Characteristics by Year.. 93

Table 4.3 – Selected Means from U.S. Mincerian Wage Regressions 94

Table 4.4 - U.S. Mean Real Hourly Wages Levels by Broad Occupation and Year........ 95

Table 4.5 – Participation Rates for Broad Occupation Categories, by year 96

Table 4.6 - U.S. Mean Real Hourly Wages Levels for Broad Industry, by year 97

Table 4.7 – Participation rates for broad industry categories, by year 98

Table 4.8 – Non-Appalachian / Appalachian Decomposition 100

Table 4.9 – Rural non-Appalachian / Appalachian Decomposition 102

Table 4.10 – Regional Demand Shifts, by year 104

CHAPTER 1

INTRODUCTION

The Appalachian region has historically experienced lower earnings than the rest of the U.S. The Appalachian mean hourly wage was nearly 18% lower than the mean non-Appalachian wage in 1940, and by 1990 the wage gap grew to 22%.[1] Labor force participation rates in 1998 were 6 percentage points lower than the national average, underscoring the lack of available jobs and discouraged workers in the region.[2] Appalachia also contains the largest concentration of white poverty in the United States. In 1990, almost 23 percent of poor white Appalachians lived in counties with white poverty rates in excess of 20 percent compared to 6.5 percent of poor whites living elsewhere in the country.[3] Although Appalachia comprises only 13 percent of all U.S. counties, it contains 30 percent of all counties with white poverty rates in excess of 20 percent. As Massey (1996) notes, the concentration of poverty is itself a social problem leading to increased exposure to crime, disease, family disruption and weaker attachment to the labor force. For all of these reasons, Appalachians are particularly disadvantaged among poor whites. Migration out of the region may raise the employment and wages of

Appalachians, but still fail to result in parity between labor market outcomes between poor white Appalachians and other poor.

There are also vast differences in the labor force in Appalachia and the rest of the country. Several studies cite lower education levels in Appalachia compared to the rest of the country (Isserman (1996), Rogers & Cushing (1996)), and this thesis finds human capital deficits for Appalachians via lower education levels and lower scores on the Armed Forces Qualifying Test. Another factor of wage growth is the effect of workforce composition. Kaboski (2002) shows that much of the wage growth experienced over the past fifty years is a combination of higher human capital levels *and* workers switching into higher paying jobs created by large increases in technology. Additional evidence of inter-industry wage differentials found in Krueger & Summers (1987 and 1998), Dickens & Katz (1987), and Gibbons & Katz (1992) illustrate the importance of industrial composition on wages.

This thesis evaluates wage differences between Appalachians and people living elsewhere in the country. The first chapter is a synopsis of the Appalachian region including the federal geographic definition, summary statistics, and a brief legislative review. The empirical analysis is broken into three parts. The second chapter compares the labor market experiences of economically disadvantaged whites (EDWs) born and raised in Appalachia with the labor market experiences of EDWs born and raised elsewhere in the country.[4] The goal is to evaluate the role played by differences in human capital, broadly defined, and differences in economic opportunities on the outcomes of these two groups. Since migration represents a potential path to better economic outcomes for many disadvantaged people, returns to migration for Appalachian and non-Appalachian EDWs are estimated. A sample of young EDW males from the National Longitudinal Survey of Youth 1979 (NLSY79) is used to compare outcomes for Appalachian EDWs with those of a nationally representative cross section of EDWs of the same birth cohort. The sample characteristics are consistent with census data: Appalachians have fewer years of education, lower levels of labor force participation, higher levels of unemployment, lower wages, lower median household income and higher poverty rates than the rest of the EDW sample. Additional evidence of disadvantage not present in Census data is found in a large Armed Forces Qualifying Test (AFQT) score deficit among Appalachians. Intended to measure an individual's trainability on the job, the AFQT score reflects the cumulative impact of unmeasured variables such as ability, school quality and family background characteristics that affect trainability.

The model consists of three equations: a reduced form discrete-time hazard model of migration, an employment equation, and a wage equation corrected for selection into employment. The model is designed to investigate (i) how much of the Appalachian wage and employment gaps seen in the raw data can be attributed to differences in human capital, (ii) whether a wage penalty exists for being born and raised an Appalachian, regardless of current location, and (iii) whether migration is a plausible solution to poverty for Appalachians. An instrumental variables estimator is used to obtain consistent estimates of migrant status in the employment and wage equations. The estimator I propose is an extension to panel data of a two step procedure commonly used with cross section data.[5] I estimate time varying probabilities of migration and then construct the predicted probability of having migrated at any previous point in time during the sample frame. I use the predicted value of migration status to obtain consistent estimates of the effects of migration on employment and wages.

I find that almost all of the 9.3 percentage point difference in the employment rates between the two groups is explained by group average differences in AFQT, education, and local economic conditions. There is no evidence of an unobserved factor common to the Appalachians holding down their employment rate. I find a 19 percent wage difference between the two groups but again there is no evidence of a common unobserved factor depressing the wages of Appalachians. A large portion of the wage differential is explained by observed differences in AFQT, education, experience, and measures of local economic conditions. The insignificance of the Appalachian native dummy variable in both the employment and wage equations belies the perception of a distinctive Appalachian subculture, a mountain or frontier spirit, of individualism that may distance Appalachians from market oriented, mainstream American society. Billings (1974) provides empirical evidence that a distinctive Appalachian culture is largely a myth. The insignificance of the Appalachian dummy coefficient, once I control for years of completed schooling and AFQT, suggests that the factors unique to the white poverty experience in Appalachia, such as the concentration of white poverty and arguably lower quality schools, affect employment and earnings only indirectly through their effects on AFQT and years of schooling.

Out-migration of Appalachians increases their probability of employment by almost 6 percentage points. Appalachians at risk of migration, i.e. non-migrants in all years and for migrants those years prior to the move, have an average employment rate of 71 percent. My results suggest that migration alone raises their employment rate to 77 percent, 3 percentage points below the employment rate of non-Appalachians at risk of migrating. My estimated migration premium for Appalachians is 7.5 percent (about $0.53 per hour evaluated at the mean wage for Appalachians at risk of migrating of $7.10), which gives Appalachians average post-migration wages that are 88 percent of the average wage for non-Appalachian EDWs at risk of migrating.[6] I find a 4 percent return to migration for other EDWs (about $0.34 per hour evaluated at the mean wage for non-Appalachians at risk of migrating of $8.69). Thus outmigration from Appalachia brings Appalachians' employment rates and wages closer in line with other EDWs, but does not eliminate the gap in labor market outcomes. The effects of lower educational attainment and lower AFQT scores cannot be overcome entirely by migration.

The third chapter uses the same empirical model and data but estimates all of the parameters simultaneously using maximum simulated likelihood (MSL). Simultaneous estimation ensures the estimates are consistent, efficient, and free of standard error bias that can result from multi-step estimation. Since the data are longitudinal, a random effect term is added to each equation to account for individual-specific unobserved variation. Correlation between the random effects is modeled, requiring a random effects covariance matrix to be estimated in addition to the model's other parameters. However, as the complexity of the model increases, so does the difficulty in both characterizing and ultimately solving the likelihood function. One tool available to facilitate solving detailed models is simulation. In this context, simulation replaces integrating each of the random effects covariance matrix parameters over the state space at each iteration.

My results show a large part of the wage and employment differences are explained by differences in human capital, measured by years of schooling and AFQT scores. I again find no evidence of an unobserved Appalachian trait that causes a wage penalty or a disassociation with the labor force. Thus, any effects on employment and wages caused by concentrated poverty or a culture of poverty are expressed through their impact on observed characteristics. Migration does not appear to be a plausible solution for poverty, especially for Appalachians. There is evidence of self-selection into migration for this sample, where the most-able candidates for migration queue into labor market segments where the returns to skill are the largest. Migration for a randomly selected Appalachian will not raise the wage to the level of a non-Appalachian EDW. This is supported by the low migration rates among Appalachians compared other EDWs raised elsewhere in the country. Since migration does not provide relief from poverty, the results suggest that the solution lies with human capital investments and development of economic infrastructure in the region. This is in contrast to the results from the second chapter, which have Appalachians earning a 7.5% premium to migration. This is a result of both the differences in estimation and the relatively small number of Appalachians observed migrating.

The fourth chapter uses a technique formulated by Oaxaca (1973) and Blinder (1973) to decompose changes in the wage gap between the Appalachia and the rest of the country into changes in quantities of observable characteristics and changes in the skill prices associated with the observables. Following Juhn, Murphy, & Pierce (1991), I further decompose the wage equation residual into changes in inequality. Data are from IPUMS (Integrated Public Use Microdata), 1940 to 1990. The observable characteristics used are education, labor market experience, black, female, industry and occupation. The goal is to break apart wage gap changes for each ten-year interval between 1940 and 1990 into contributions from changes in the relative quantities of observable characteristics, skill prices, and wage equation residuals. My results show the decade where the wage gap increased the most is the 1980s. During both of these decades changes in skill prices and an increase in wage inequality were favorable to non-Appalachians. Non-Appalachians also benefited from quantity changes in industries and occupations in every decade except the 1940s. Larger increases in participation rates in higher paying industries and occupations by non-Appalachians is likely related to higher levels of human capital and a larger stock of high paying jobs.

A potential cause of the wage gap between Appalachia and the rest of the country is differences in urban-rural composition. An additional sample of rural non-Appalachians was drawn to test if urban-rural composition is driving the results. Mean weekly wages between Appalachians and rural non-Appalachians in 1990 differ by less than two percent, suggesting much of the wage gap between Appalachia and the rest of the country can be attributed to differences in urban-rural composition. The wage gap between these groups has fluctuated over the sample frame. Large wage growth for rural non-Appalachians during the 1940s and 1950s put rural non-Appalachians ahead of

Appalachians until the 1970s, when a 10% dip in mean hourly wages for rural non-Appalachians leveled wages between the two groups.

I use the same data set to estimate labor demand indexes formulated by Murphy & Katz (1992) for each region to test for effects of national industry and occupation trends on each of the three regions. The Appalachian labor demand change is negative and the non-Appalachian demand change is positive for each decade in the sample except the 1940s. Further investigation into the industry and occupation shares show that the largest Appalachian industries and occupations (e.g., manufacturing) have experienced large national declines in demand. Demand shifts towards those with higher skills have been largely unfavorable to the below-average skills of the Appalachian workforce.

1.2 Overview of Economic Conditions in Appalachia

The Appalachian region is federally defined as 410 counties in parts of 12 states— Alabama, Georgia, Kentucky, Maryland, Mississippi, Ohio, New York, North Carolina, Pennsylvania, South Carolina, Tennessee, and Virginia— and all of West Virginia. This region had a total population of 20,701,881 in 1990. Whites constitute 92.2 percent of the population. Blacks constitute 7 percent of the population,[7] with most Appalachian blacks living in Mississippi.[8] Historically, Appalachia has lagged behind the rest of the country in terms of human capital, economic opportunity, income, and employment. The region has made many economic advances since the 1960s, but it continues to trail behind the rest of the country in many economic indicators.

The economic problems of Appalachia first came to national attention in 1964 when President Kennedy formed the President's Appalachian Regional Commission

(PARC). This federal program was established in response to the low levels of income, education, and employment in the region (Isserman (1997)). One of its principal accomplishments was the establishment of the Appalachian Regional Commission (ARC), which has overseen policy for the area ever since. The ARC is comprised of the thirteen governors from the Appalachian region and a presidential appointee. Annually the thirteen governors propose a funding plan for the year for their respective states, and the ARC reviews each proposal and decides how to spend federal funding appropriated by Congress. Congress has appropriated $7.8 billion for Appalachian development programs through September 30, 1999.[9]

The original PARC report described Appalachia's poverty as a problem of isolation. One of the original goals, which remains a top priority today, is the development of Appalachian infrastructure. To combat the problem of isolation, the Appalachian Development Highway System (ADHS) was devised by the ARC in 1965. The ADHS is a 3,025-mile roadway system that creates new roads and improves existing ones in order to connect Appalachia with the rest of the country. By September 2000, 2,483 miles (approximately 82 percent) were complete or under construction, with the remaining 542 thought to be the most expensive to build. Of the $7.8 billion appropriated to the ARC by Congress since 1965, $5.0 billion went to ADHS.

In the 1960s, Appalachia dramatically lagged behind the rest of the United States in many economic indicators. The PARC report described Appalachia as "deeply unemployed," citing the 6.8 percent unemployment rate in Appalachia compared with the national rate of 5.0 percent in 1960. Not surprisingly, in 1969 the Appalachian per capita income was 78 percent of the rest of the country. Over the past thirty years, the Appalachian region has seen sharp increases and decreases in growth.[10] Despite these fluctuations, Appalachia is less entrenched in poverty than it was before. However, these economic gains within Appalachia have not brought the region even with the rest of the nation. For example, in 1990 the number of high school graduates was in Appalachia was 7 percentage points less than the rest of the country (Isserman (1996)).

Table 1.1 presents statistics from the 1990 Census that summarize white disadvantage. Since Appalachia is less urbanized than the rest of the country it is useful to divide the summary statistics by metropolitan and non-metropolitan areas. White poverty rates are almost 50 percent higher in Appalachia than the rest of the country. Comparison of white poverty rates between urban and rural areas yield similar differences. Moreover, white poverty tends to be highly concentrated within the region. Appalachian counties represent approximately 13 percent of all the counties in the U.S., but the region included 29 percent of the counties with white poverty greater than 20 percent in 1990. Most of these counties are in southern Ohio, eastern Kentucky and West Virginia.[11]

Table 1.1 also shows that in 1990 unemployment rates among whites remained higher in Appalachia than the rest of the country. The white unemployment rates stood at 7.77 percent in non-metropolitan Appalachia, about 31 percent higher than the rest of the non-metropolitan U.S. In metropolitan areas, white unemployment rates were 13 percent higher in Appalachia than elsewhere. Mean white household income of $30,534 in

Appalachia was only 78 percent of mean white household income in the rest of the country, the same as in 1969.

Education deficits also exist among Appalachians. In 1990, among people of all races aged 25 and older, the high school dropout rate was almost 32 percent in Appalachia, compared to 24 percent in the rest of the country. The fraction of the population with a terminal high school degree was almost 35 percent in Appalachia, compared to 28 percent elsewhere. Predictably, the fractions with some college or a college degree were commensurately lower in Appalachia as well. The relatively rural environment of the region explains some of the Appalachian educational deficit, since people residing in metropolitan areas tend to be better educated than those living in rural areas.

	Appalachia	Rest of the United States
Percentage of Whites Living Below Poverty Level, 1990		
Total	13.5 %	9.1 %
Nonmetro	17.2 %	12.9 %
Metro	10.6 %	8.1 %
Total White Male Unemployment, age greater than 16, 1990		
Total	6.56 %	5.22 %
Nonmetro	7.78 %	5.92 %
Metro	5.73 %	5.06 %
Mean White Household Income, 1990		
Total	$30,534	$39,191
Nonmetro	$26,761	$29,024
Metro	$33,166	$41,509
Educational Attainment age 25 and older, all races, 1990		
Total		
Less than HS grad	31.63 %	24.13 %
HS grad	35.04 %	29.52 %
Some college	19.08 %	25.45 %
College degree	14.26 %	20.90%
Nonmetro		
Less than HS grad	37.66 %	29.82 %
HS grad	35.34 %	34.79 %
Some college	16.34 %	22.21 %
College degree	10.66 %	13.19 %
Metro		
Less than HS grad	27.43 %	22.85 %
HS grad	34.83 %	28.33 %
Some college	20.98 %	26.18 %
College degree	16.76 %	22.64 %
Source: White poverty statistics are author's calculations from the 1990 Census. The remaining statistics come from Rogers & Cushing 1996a & 1996b based on the 1990 Census.		

Table 1.1 - Poverty, Unemployment, Income and Education Statistics

CHAPTER 2

A COMPARISON OF EMPLOYMENT, EARNINGS, AND MIGRATION OF ECONOMICALLY DISADVANTAGED WHITES IN APPALACHIA AND THE REST OF THE UNITED STATES

This section employs a comparison of economically disadvantaged whites (EDWs) born and raised in Appalachia and elsewhere in the country. I chose to use a sample comprised entirely of EDWs to illustrate relative disadvantage to those born and raised in the Appalachian region. The sample means presented in this paper show that even amongst a sample of EDW Appalachians have deficits in human capital, earnings, and employment rates. Thus, the Appalachian region represents a concentrated area of white poverty, which potentially leads to increased exposure adverse societal outcomes such as to crime, disease and family disruption (Massey (1996)).

The model is designed to investigate (i) how much of the Appalachian wage and employment gaps seen in the raw data can be attributed to differences in human capital, (ii) whether a wage penalty exists for being born and raised an Appalachian, regardless of current location, and (iii) whether migration is a plausible solution to poverty for

Appalachians. These questions imply three endogenous regressors: wages, employment, and migration. A multi-step estimator is used to obtain consistent results. My results show that the gaps in the employment rate and earnings between EDW Appalachians and non-Appalachians is explained by observed characteristics, including human capital proxies and local economic conditions. After this introduction, this chapter is broken into four parts: model, data, results, and conclusions.

2.2 Model

The observation period on each individual in the sample begins with the year that the individual first leaves school for at least twelve months. I follow each individual until he moves a second time, misses an interview, or reaches the end of the sample frame with continuous interviews. The model allows at most one move. I do not consider return migration or migration to another labor market after the initial move.

The model consists of three equations: a reduced form discrete-time hazard model of migration, an employment equation and a wage equation corrected for selection into employment. There is non-zero correlation between the error terms in all three equations. The correlations between the three error terms imply that migrant status, which enters the employment and wage equations, is endogenous in both equations. Using the multi-stage model gives consistent estimates of the effect of migration on employment and wages.

The reduced form migration equation includes all the variables in the employment equation, all the variables in the wage equation except job-specific variables, plus variables, excluded from the employment and wage equations, that subsequently identify migration in the other two equations. Let m_{ijt} be a latent variable representing the net present value of moving to the best alternative labor market for individual *i* still living in original labor market *j* at time $t-1$. Let X_{ijt} represent a vector of all individual and local labor market characteristics that affect the reduced form propensity to migrate,

$$m_{ijt} = \beta' X_{ijt} + \varepsilon_{ijt}, \tag{1}$$

$$\text{where } \varepsilon_{ijt} = \eta_{ijt} + \nu_i .$$

The error term, ε_{ijt}, is comprised of the sum of an individual-specific error term ν_i, (i.e., the random effect) and an overall error term η_{ijt}. I assume that both are normally distributed with $\eta_{ijt} \sim N(0,1)$, $\nu_i \sim N(0,\sigma^{\nu})$ and $E(\eta_{ijt}\nu_i) = 0$. The inclusion of the random effect term implies that the unobserved propensity to migrate is correlated over time for each individual. Furthermore, the assumption that the random effect is distributed normally throughout the population implies that the unobserved time invariant individual specific propensity to migrate is unimodal and continuously distributed. I control for duration dependence in the propensity to migrate by including a variable representing the years spent in the original labor market in vector X_{ijt}.

Let M_{ijt} be a dichotomous choice variable for individual *i* indicating the decision to move out of local labor market *j* at time *t*, where

$$M_{ijt} = 1 \text{ if } \eta_{ijt} > -\beta' X_{ijt} - \nu_i \text{ and } M_{ijt} = 0 \text{ if } \eta_{ijt} \le -\beta' X_{ijt} - \nu_i .$$

The discrete time hazard of migration is the probability that i moves at time *t* conditional on remaining in j until period $t-1$ is

$$\Pr(M_{ijt} = 1 \mid M_{ijt'} = 0 \ \forall \text{ t'<t, } \mathrm{X}_{\mathrm{ijt}},\ \nu_{\mathrm{i}}) = \Phi(\beta' X_{ijt} + \nu_i), \tag{2}$$

where Φ is the standard normal cumulative distribution function. To evaluate this probability at the mean of the random effect, I need to integrate out the random effect as follows

$$\Pr(M_{ijt} = 1 \mid M_{ijt'} = 0 \ \forall \ \text{t'<t}, \ \mathrm{X_{ijt}}) = \int_{-\infty}^{\infty} \Phi(\beta' X_{ijt} + \nu_i)\varphi(\nu_i)d\nu_i \ , \qquad (2')$$

where $\varphi(\nu_i)$ is a zero mean normal density with standard deviation σ^{ν}.

The likelihood of migrating in period t is the joint probability of migrating in t and not migrating in any period prior to t. If the error terms in the discrete time hazard, $\eta_{ijt} + \nu_i$, are independent over time, the hazards would be independent over time. The likelihood of migrating in period t is obtained by multiplying the hazard of migrating in t by the product of the hazards of not migrating in any period $\tau < t$. However, the individual's errors are not independent over time because of the time-invariant, individual specific random effect, ν_i.

Although the individual's hazards of migration are not time independent, the expected values of the individual's hazards, where expectations are taken over the state space of the random effect, are time independent. The random effect, ν_i, is integrated over its state space to create an expected value of the predicted hazard of migration for each person in each period. An alternative method relies on the law of large numbers and achieves the same result by simulating the random effect and averaging the predicted probability of migration across simulations.

The deterministic part of the hazard can be predicted for each individual in each time period from estimates of $(\hat{\beta}, \hat{\sigma}_v)$ and the vector X_{ijt}. To implement the simulation of the expected value of the hazard of migration, 1000 random draws of v_r are drawn from the distribution $N(0, \hat{\sigma}_v)$. Then, holding constant v_r across individuals, a vector of predicted discrete time hazards is computed for each individual i in each of the T_i periods that i is observed in the sample frame[12]

$$\left(\Phi(\hat{\beta}' X_{ij1} + v_r), \ldots, \Phi(\hat{\beta}' X_{ijT_i} + v_r) \right). \tag{3}$$

The average over the 1000 draws of v_r is taken to purge the predicted migration hazards of the time invariant random effect. Using the law of large numbers,

$$\hat{\Gamma}_{ijt} = E_{v_r} \hat{\Pr}(M_{ijt} = 1 \mid M_{ij\tau} = 0 \ \forall \ \tau\text{<t}) =$$

$$\int_{-\infty}^{\infty} \Phi\left(\hat{\beta}' X_{ijt} + v_r \right) \varphi(v_r) dv_r \approx \frac{1}{1000} \sum_{r=1}^{1000} \Phi(\hat{\beta}' X_{ijt} + v_r), \tag{4}$$

where r indicates the r^{th} draw of the random effect in each round of the simulation and $\varphi()$ is a normal density function with mean zero and standard deviation $\hat{\sigma}_v$.

The expected value of the predicted likelihood that i moves in period t and does not move in any period $\tau < t$ is

$$E_{v_r} \hat{\Pr}(M_{ijt} = 1 \ \& \ \mathrm{M}_{\mathrm{ij}\tau} = 0 \ \forall \ \tau < t) =$$

$$E_{v_i} \hat{\Pr}(M_{ijt} = 1 \mid M_{ij\tau} = 0 \ \forall \ \tau < t \ ) \prod_{\tau < t} E_{v_i} \hat{\Pr}(M_{ij\tau} = 0 \mid M_{ij\tau'} = 0 \ \forall \ \tau' < \tau \ ) =$$

$$\hat{\Gamma}_{ijt} \prod_{\tau<t}\left(1-\hat{\Gamma}_{ij\tau}\right). \tag{5}$$

The likelihood is the joint probability that migration occurs in period t and does not occur in any $\tau < t$.

The expected value of the cumulative likelihood that i migrates out of labor market j by period t is the sum of the expected value of the likelihood that i moves in any period prior to and including t. To formalize this notion, let migrant status be represented by an indicator variable, I_{ijt}, that takes a value of 1 if the individual has already migrated out of original labor market j at time t and 0 otherwise. The expected value of the predicted cumulative likelihood of migration for each individual i in each of the periods that i is observed is

$$\hat{I}_{ijt} = E_{v_i} \hat{\Pr}(I_{ijt}=1) = \sum_{\tau'=1}^{t} \hat{\Gamma}_{ij\tau'} \prod_{\tau<\tau'}\left(1-\hat{\Gamma}_{ij\tau}\right). \tag{6}$$

Observed migrant status, I_{ijt}, is correlated with the error terms in the wage and employment equation because the unobservables driving the migration decision are correlated with the unobservables driving wages and employment. Using observed migrant status as a regressor in wage and employment equations will produce biased estimates of the effects of migrant status on these two variables of interest. However, $\hat{I}_{ijt}$ is orthogonal to errors in all three equations by construction. The multi-stage estimation model that I propose uses the expected value of the predicted cumulative likelihood of migration, $\hat{I}_{ijt}$, instead of observed migrant status I_{ijt} to obtain consistent estimates of the effect of migrant status on wages and employment.

Specification of the wage and employment equations in the model begins with defining the participation rule. Let w_{ijt}^{res} denote the reservation wage for individual *i* in period *t*.[13] Let Z_{ijt} represent a vector of individual and local labor market characteristics that affect the reservation wage. The reservation wage equation is written as

$$w_{ijt}^{res} = \alpha' Z_{ijt} + \xi_{ijt}, \tag{7}$$

$$\text{where } \xi_{ijt} \sim N(0, \sigma^{\xi}).$$

Let w_{ijt} be the wage offered to individual *i* in market *j* at time *t*. Let Y_{ijt} represent a vector of individual and local labor market characteristics that affect the wage offer. The wage offer equation is written as

$$w_{ijt} = \gamma' Y_{ijt} + \phi I_{ijt} + \mu_{ijt}, \tag{8}$$

$$\text{where } \mu_{ijt} = \varpi_i + \varsigma_{ijt}.$$

I assume $\varpi_i \sim N(0, \sigma^{\varpi})$ and $\varsigma_{ijt} \sim N(0,1)$.

If the endogenous migrant status, I_{ijt} (which is correlated with error term μ_{ijt}), is replaced with $\hat{I}_{ijt}$, (which is uncorrelated with μ_{ijt} by construction), the quasi-reduced form employment choice equation becomes

$$E_{ijt} = 1 \text{ if } \varsigma_{ijt} - \xi_{ijt} > \alpha' Z_{ijt} - \gamma' Y_{ijt} - \phi \hat{I}_{ijt} - \varpi_i, \text{ and } E_{ijt} = 0 \text{ otherwise.}$$

Let $\sigma^{\varsigma-\xi} = \sqrt[1/2]{1+\sigma^{\xi^2} - 2E(\varsigma\xi)}$ and H_{ijt} be a vector containing all elements of Z_{ijt}, $-Y_{ijt}$ and $-\hat{I}_{ijt}$. Furthermore, let H^*_{ijt}, κ^*, and ϖ^*_i and be the vectors H_{ijt} and κ and the random effect ϖ_i, respectively, normalized by $\sigma^{\varsigma-\xi}$. The quasi-reduced form probability of employment is

$$\Pr(E_{ijt} = 1) = 1 - \Phi\left(\kappa^* H^*_{ijt} + \varpi^*_i\right). \tag{9}$$

Equation (9) is estimated with a random effects probit. I am only able to estimate the variance of the individual specific random effect relative to the variance of $\varsigma - \xi$. To obtain estimates of the wage offer equation, equation (8), from the subsample of men observed working, I use the Heckman-Lee two-step selection procedure (Lee (1978), Heckman (1979)). Estimates from the employment probit are used to construct the inverse Mills ratio, which is then included as a regressor in the wage equation to correct for sample selection. I do not condition on the random effect in constructing the inverse Mills ratio, which implies that the random effect in the selection corrected wage equation is proportional to the random effect in the wage offer equation. The structural wage equation that I estimate is

$$w_{ijt} = \gamma' Y_{ijt} + \phi \hat{I}_{ijt} + \theta \hat{\lambda}_{ijt} + \varpi^{**}_i + \xi_{ijt}, \tag{10}$$

where $\hat{I}_{ijt}$ is the predicted probability of having left the original labor market, $\hat{\lambda}_{ijt}$ is the predicted inverse Mills ratio constructed from the quasi-reduced form employment equation, and ϖ^{**}_i is an individual specific effect. As Heckman (1979) and Lee (1978) demonstrate the coefficient estimates from this kind of two-step estimator are consistent. The wage equation is estimated with random effects.

2.3 Data

The data used in this analysis come from the 1979-1990 waves of the National Longitudinal Surveys of Youth 1979 (NLSY79).[14] Respondents, born between 1957 and 1964, are representative of this birth cohort living in the United States in 1978. The NLSY79 sample design called for a nationally representative cross section, oversamples of African Americans, Hispanics and poor whites, and a subsample serving in the military. I restrict attention to male sample members who were living in poverty at least once between 1978 and age 21. This definition of economic disadvantage is an extension of the one used by the NLS that categorized poor whites as having lived in poverty in 1978. The strength of the NLSY79 in studying labor market experiences of poor whites lies in the level of detail the data have to offer and repeat observations on the same individuals. The NLSY79 has data on AFQT scores and years of education, which provide good measure of human capital.[15] It also has data on county of birth and county of residence at age 14, which used to define native Appalachians.

Table 2.1 details how the sample was constructed. There are 2,770 economically disadvantaged males in the NLSY79. Eighty individuals who were born in the U.S. but had an unknown state or county of birth were deleted. Forty-two additional respondents were deleted who had an unknown state or county of residence at age 14. I also deleted 182 immigrants who were born abroad to parents who were not U.S. citizens. This leaves

324 Appalachians and 2,142 non-Appalachians. I define someone as being Appalachian if they were born in Appalachia and resided there at age 14.

Since my focus is on non-Hispanic white poverty, the next deletions came from self-reported ethnicity. Ninety-nine American Indians, 227 Hispanic, 11 Asians, and 771 African Americans were deleted. Sixty people whom the interviewer identified as non-white in the original 1979 interview were deleted. Since I am interested in the employment decisions of individuals after leaving school, 34 people for whom the date they first left school for a minimum period of 12 months could not be identified were deleted. Finally identifying migration of the respondent requires at least two consecutive civilian interviews, resulting in 85 people deleted. The final sample consists of 193 Appalachians and 986 non-Appalachians, for a total of 1,179 EDW males. Respondents continuously interviewed since leaving school are followed until they either are first observed to move, first miss an interview or reach the end of the sampling frame, whichever comes first. The end of the sample frame is 1990 because that is the last year in which the oversample of poor whites were interviewed.

I use the full sample to estimate the probability of employment. The migration equation is estimated on the sample of all people at risk of migrating. The wage equation is estimated on the subsample of men who were employed at the survey date throughout the sample frame, and so potentially includes multiple post migration observations for each migrant. Means for the variables used in each of these three subsamples are presented.

Table 2.2 presents the means for the variables used in the migration equation. For non-Appalachians I define a move as a change in MSA or changing counties that have centroids at least 75 miles apart.[16] All moves by Appalachians require a move outside of Appalachia to be counted as a move. These observations are pooled over time and across individuals into person years. There are 1,342 person/years for the Appalachian respondents and 4,918 person/years for the non-Appalachian respondents.

Among the Appalachians, 25.18 percent migrate whereas 37.22 percent of the non-Appalachians migrate.[17] Appalachians have almost 1.5 fewer years of schooling than non-Appalachians and score 13.3 percentage points lower on AFQT tests. There is controversy about whether AFQT measures ability or more general job readiness. However, the low AFQT scores among Appalachians suggest that they are hard to train and hence less able to accumulate human capital on the job. This difference may reflect lower school quality (Hanushek (1973)), the cumulative effect of concentrated poverty and/or disadvantage resulting from family background (Neal and Johnson (1996)).

To capture the local economic conditions of the labor market the respondent is considering leaving, I use previous year county unemployment rate and county retail earnings. This decision is motivated in part because the precise date of migration is not known, only that the migration occurred at some point between two interviews. Mean lagged county unemployment is 1.8 percentage points higher for Appalachian natives and lagged weekly county retail earnings are lower by almost 16 dollars.[18] I also include county median rent to reflect cost of living differences not otherwise captured by county retail earnings.

The variables included in the migration equation, but excluded from the employment and wage equations, are the instruments that identify migration in the three stage estimates. These variables are dummy variables for whether the respondent got married or divorced between the two interviews.

Table 2.3 presents means for the variables used the employment equations. Any person years that identified schooling as the primary activity during the survey week were removed. This produced 1,576 Appalachian person/years and 6,545 non-Appalachian person/years. Seventy-two percent of the Appalachians were employed during the survey week, compared to 81 percent of the non-Appalachians, with a statistically significant difference in means. Eight percent and 16 percent of the person/year observations are post-migration for the respective Appalachian and non-Appalachian subsamples.

The person/year human capital variables reflect the Appalachian/non-Appalachian differences that were seen Table 2.3. About 4 percent of person/years reflect health limitations that limit the amount or kind of work the respondents can do. Around 39 percent of the Appalachian sample and 36 percent of the non-Appalachian sample involve years when the respondents are married with spouse present. Around 55 percent of the Appalachian sample and 69 percent of the non-Appalachian sample involve years when the respondents are living in a county designated as a metropolitan area. There are also statistically significant differences in the means of county unemployment rates (2 percentage points), lagged county unemployment rates and retail earning (2 percentage points), and county weekly retail earnings ($23) which proxy for local labor market conditions.[19] This reflects the relative lack of economic opportunities within Appalachia, where 91 percent of the Appalachian person/year sample are observed.

Table 2.4 presents the person/year means for the variables used in the wage equation. To be included in the sample, the respondent had to be employed during the survey week, report a wage between $1 and $250 per hour in 1990 dollars, and have valid responses to all other variables. This produced 1,202 Appalachian person/years and 5,392 non-Appalachian person/years. The most striking figure is the statistically significant difference in real wages. Appalachians earn $1.72 per hour less than non-Appalachians, or 80 percent of the non-Appalachian average wage. Some of this wage difference is attributable to differences in observed characteristics, such as lower levels of education, AFQT, less felicitous local economic conditions and lower cost of living. Since experience is measured as age minus education minus six, Appalachians have higher mean potential experience, reflecting the difference in years of schooling. However, there is no statistically significant difference in tenure, suggesting that the problem facing Appalachians is finding work, not holding onto a job once it has been acquired.

Appalachians who find work are 13 percent less likely than non-Appalachians to live in an urban area. The difference in county unemployment rates is 1.8 percentage points and the difference in county retail earnings is $23. Overall, the means underscore what has been said before: Appalachians live in areas where economic opportunities are less advantageous than for other EDWs.

I include median rent in the wage equation to proxy for cost of living differences not otherwise captured by county retail earnings. The two variables are not perfectly correlated. The Appalachian non-Appalachian gap in retail earnings is on the order of 90 percent, whereas the gap in median rent is 80 percent. Although median rent is not a standard variable in wage regressions, it is included to purge cost of living differences from the return to migration.

The final variable used in the wage model is an inverse Mills ratio, calculated for each person from the estimates of the employment probit. The variables used to identify the selection correction for employment in the wage equation are lagged county unemployment rate, age and the percent of the population residing in an urban area in the respondent's county of birth during either 1956 or 1962, whichever is closest to the respondent's birth year.[20]

2.4 Results

The migration equation is estimated using a probit model with random effects. The results are presented at Table 2.5. Getting married, one of the variables excluded from the other two equations, increases the probability of migration by over 2.7 percentage points, while an increase in retail earnings by $100 per week decreases the probability of migration by almost 1.6 percentage points. The remaining variables have the expected sign, but none has a marginal effect in excess of one percentage point on the probability of migration.

The other variable used to identify predicted migrant status in the remaining two equations, get divorced, has a statistically insignificant effect on the probability to migrate. However, the large and statistically significant effect of the get married variable provides evidence that weak instruments are not an issue and predicted migration status is identified.

Table 2.6 presents the results of the employment model estimated using a random effects probit. Columns two and three report the results using actual migrant status, which I argue is endogenous and correlated with the error term in the employment equation. Columns three and four report results using the predicted value of migrant status, which is uncorrelated with the error term in the employment equation by construction. Migrant status has a statistically significant impact on employment for Appalachians using the multi-step estimator.

Migration causes an almost 6 percentage point increase in the probability of employment for Appalachians. However, migration has no statistically significant effect on employment of non-Appalachians. These results are striking because they arise in the instrumental variables estimator and so cannot be explained by the self-selection of more highly motivated migrants. I suspect that the variables for local economic conditions do not fully capture the difference in job opportunities and job quality inside and outside of Appalachia. This is bolstered by the finding that Appalachian migrants receive a greater wage boost than non-Appalachian migrants. I defer a more detailed discussion until after results from the wage equation are presented.

Another interesting result is the insignificant coefficient on the Appalachian native dummy. This suggests that the employment effects arising from a culture of poverty begotten by the high concentration of white poverty in the region are either absent or reflected in observed characteristics, in particular lower years of schooling and AFQT, that retard employment. The nine percentage point difference in employment found in the sample means can be attributed to differences in observed characteristics, rather than an unobserved trait common to Appalachians.

Of the other regressors in the model, married spouse present and experience have a large positive effects, increasing the probability of employment by 4.3 and 2.9 percentage points, respectively. The proxies for the local conditions suggest that individuals residing in counties with stronger economies are more likely to be employed. Living in a metropolitan county has no significant effect on employment, but a one percentage point increase in the county unemployment rate decreases employment by roughly 0.7 percentage points. Also, increasing county retail wages by $100 per week increases the probability of employment by over 4 percentage points.

Table 2.7 presents the estimates for the wage equation. The dependent variable is the natural log of real hourly wages in 1990 dollars. A random effects estimator is again used to correct for time invariant unobserved individual effects arising from the longitudinal data. Sample selection into employment is corrected by a Heckman selection correction term, where the inverse Mills ratio is created from estimates of the employment probit reported in Table 2.6 evaluated for each individual at the mean random effect of zero.

Two specifications of the wage model are presented- one with the predicted migrant status and one with endogenous actual migrant status. All the coefficients, except for those on migrant status, are very similar across the two specifications. The coefficient on the Appalachian dummy variable is insignificant. There is no evidence that

Appalachian natives earn less for some common unobserved reason. Differences in human capital and local economic conditions, however, explain a large part of of the 19 percent wage differential than the Appalachian dummy variable. AFQT and highest grade completed are both positive and statistically significant in the log wage equations. An additional year of schooling increases wages by over 7 percent, and an increase of one percentage point in the AFQT score leads to a 0.5 percent increase in wages. Tenure and total experience also have the expected signs with the familiar concave effect on wages. The unemployment rate and retail earnings are also significant and of the expected signs. Median rent, an additional control for cost of living differences is positive and statistically significant. The selection correction estimate, lambda, is positive and statistically significant, suggesting a positive correlation between the error term in the employment probit and the wage equation. People with unobserved characteristics that lead to higher earnings are more likely to be working in this sample.

Turning to estimates of the migration premium, the results are sensitive to use of endogenous migrant status versus predicted migrant status. When actual migrant status is used, Appalachians earn a 8.5 percent return to migration, while non-Appalachians migrants earn just over 2.2 percent, although the latter is not statistically significant. The second column of estimates uses the predicted probability of migration interacted with the Appalachian native dummy. Appalachians earn a 7.5 percent premium to migration while non-Appalachians earn a 4 percent premium. Both are statistically significant. The estimated returns to migration in column two, unlike the ones reported in column one, represent the return to migration for a randomly selected individual. In the raw data the real hourly wages of Appalachians and non-Appalachians at risk of migration are $7.10 and $8.69, respectively.[21] Thus, controlling for differences in observable characteristics, Appalachian migrants earn an additional $0.53 per hour and non-Appalachian migrants earn an additional $0.34 per hour, measured in real 1990 dollars.

There are a number of reasons why Appalachian EDWs earn higher returns to migration than non-Appalachian EDWs. The following discussion also relates to my finding that migration raises the probability of employment for Appalachians but not for non-Appalachians. First, Appalachia is a region with a relatively stagnant economy. Blanchard and Katz (1992) demonstrate that states with counties in Appalachia have experienced relatively low rates of employment growth over the past fifty years. This in itself can lead to fewer good jobs in the region than outside. Second, Weiler (2001) shows that the West Virginian labor market, which is representative of large parts of Appalachia, is characterized by a dual labor market with a small unionized sector that pays relatively well and a much larger low wage secondary sector. Dual labor markets create incentives for non-employment since workers often find it in their best interest to queue for jobs in the high wage sector rather than accept a job in the secondary sector. I do not want to suggest that dual labor markets do not exist outside of Appalachia, but the evidence suggests that labor markets are more highly segmented in Appalachia than elsewhere. Since I define migration for Appalachians as a move to someplace outside Appalachia, Appalachians are moving from an area with relatively few "good" jobs to areas with better job prospects. Other EDWs are for the most part migrating across counties outside Appalachia. They, too, are likely to relocate to areas with better job opportunities than their original location but the gap between the areas they are leaving and the areas to which they are moving is likely to be smaller that that experienced by Appalachian migrants.

Although local economic opportunity is controlled for directly in the wage equation through county unemployment rate and weekly retail earnings variables, there are likely to be unmeasured differences in job opportunities that influence the magnitudes of the coefficients on the returns to migration for the two groups.[22] For example, it is well understood that official unemployment figures mask the discouraged worker effect. People, particularly in areas of high unemployment, are known to become so discouraged that they stop actively seeking employment and disappear from both the numerator and denominator of the unemployment rate. Unmeasured differences in local economic conditions between Appalachia and the remainder of the country could explain why Appalachians who emigrate from the region have a large measured increase in employment whereas non-Appalachian migrants, moving for the most part between two labor markets outside of Appalachia, do not. It also explains why Appalachian migrants experience larger wage gains than non-Appalachian migrants.

Part of the reason that Appalachians experience a larger wage boost than non-Appalachians when they migrate may be unmeasured cost of living differences between Appalachia and the rest of the country. I have tried to control for this with county retail earnings and median rent. But I cannot rule out that unmeasured cost of living differences contribute to the higher migration premium for Appalachians.

2.5 Conclusions

Appalachia is unique region, relatively isolated and marked by historically high unemployment rates. In 1990, white poverty rates in Appalachia were 50 percent higher than in the remainder of the country. White poverty in Appalachia was also more concentrated than elsewhere. Almost 23 percent of poor white Appalachians lived in counties with white poverty rates in excess of 20 percent. Only 6.5 percent of poor whites in the remainder of the country lived in counties with comparably high white poverty rates. Concentrated poverty can itself breed social ills, including weak attachment to the labor force (Massey (1996)), raising the question whether Appalachian whites are particularly disadvantaged relative to other poor white Americans.

This paper compares economic outcomes of economically disadvantaged white Appalachian males with those of economically disadvantaged white males born or raised elsewhere in the country. It develops a longitudinal model of migration, employment and wages to assess the factors that explain differences in employment and wages between the two groups and to quantify the effects of migration on labor market outcomes of the two groups. Data from the National Longitudinal Surveys of Youth 1979 is used to estimate the model.

In the raw data, the employment rate of poor white Appalachians is 9.3 percentage points lower than other EDW men and their wage is 19 percent lower. A large part of these differences is explained by differences in human capital, measured by years of schooling and AFQT scores, a measure of an individual's trainability on the job. Measured regional differences in economic opportunity are also important determinants of the relatively poor economic performance of Appalachian men. Higher unemployment rates and lower weekly earnings in retailing inside Appalachia contribute substantially to Appalachian employment and wage deficits. I find no evidence that unmeasured characteristics associated with growing up in Appalachia adversely affect labor market outcomes. Thus, any effects on employment and wages of concentrated poverty or a culture of poverty are expressed through their impact on observed characteristics such as years of schooling and AFQT score.

Migration out of Appalachia may ameliorate the effects of regional economic deprivation. I find that migration raises the probability of employment for Appalachians by almost 6 percentage points. No effect of migration on employment is found for non-Appalachians. I find positive and statistically significant returns to out-migration for poor white Appalachians and smaller but statistically significant returns for non-Appalachian migrants. Among Appalachians, the estimated return to migrating out of the region is 7.5 percent, whereas the estimated return for migration among non-Appalachians is 4 percent.

These results suggest that the long-term solution to the persistent high concentrations of white poverty in Appalachia lie in improving both the quantity and quality of education and developing economic opportunity within Appalachia. Migration out of Appalachia increases both employment and wages, but it does not entirely eliminate the disadvantage that Appalachian natives experience in the labor market.

	Appalachian Born and Living at Age 14 in Appalachia	Non-Appalachian Born or Living at Age 14 Outside of Appalachia	Total Number of Individuals
NLS data- male poor white oversample or economically disadvantaged white cross section males	–	–	2770
After deleting respondents born in U.S. with unknown state or county of birth	–	–	2690
After deleting respondents born in U.S. with unknown state or county of residence at age 14	325	2323	2648
After deleting born abroad to non-U.S residents	324	2142	2466
After deleting self-reported American Indian ethnicity	297	2070	2367
After deleting self-reported Hispanic ethnicity	294	1846	2140
After deleting self-reported Asian ethnicity	293	1836	2129
After deleting self-reported African ethnicity	212	1146	1358
After deleting individuals the interviewer identified as non-white	211	1087	1298
After deleting people for whom the date first left school for at least 12 months (wkout12) could not be identified	211	1078	1289
After deleting people observed only on or before wkout12	207	1057	1264
After deleting people with less than 2 consecutive interviews as civilians	193	986	1179

Source: National Longitudinal Surveys of Youth 1979.
The NLSY79 followed the 1978 poverty guidelines for farm and non-farm residents in determining who is economically disadvantaged. A non-farm resident is economically disadvantaged if family income was less than \$3140 + 1020*(number of family members – 1) in 1978. A farm resident is economically disadvantaged if family income was less than \$2690 + 860*(number of family members – 1) in 1978. The farm population lives in a rural area and either has a place with less than 10 acres and sales of farm products of at least \$250 or lives on a place with more than 10 acres and had sales of farm products of at least \$50. A respondent is designated economically disadvantaged based only on these criteria in the first interview. The designation does not change year to year.

Table 2.1 - Sample Selection Criteria

	Total	Appalachian	Non-Appalachians	Difference in Means [t-test]
Ever Migrate****		0.2518	0.3722	
AFQT**	-7.5054 (23.0800)	-17.8895 (21.0410)	-4.5686 (22.7835)	-13.3209 [19.2004]
AFQT Missing	0.0310	0.0022	0.0388	-0.0366 [6.883]
Highest Grade Completed	11.7840 (2.536)	10.6423 (2.2988)	12.0956 (2.5080)	-1.4523 [19.1460]
Married Spouse Present	0.3996	0.4188	0.3943	0.0245 [1.625]
Get Married Since Last Interview	0.0727	0.0715	0.0768	-0.0052 [0.6474]
Get Divorced Since Last Interview	0.0256	0.0261	0.0254	0.0007 [0.1365]
Health Limitations	0.0387	0.0432	0.0374	0.0058 [0.9760]
Metro*	0.6487	0.5529	0.6749	-0.1220 [8.3403]
Lagged Unemployment Rate*	7.8180 (3.5096)	9.2994 (3.9213)	7.4138 (3.2746)	1.8856 [17.8842]
Two Period Lagged Unemployment Rate*	7.4080 (3.1617)	8.4405 (3.4403)	7.1263 (3.0209)	1.3142 [13.6966]
Lagged Weekly Retail Earnings (in $100, 1990$)*	2.8846 (0.4408)	2.7615 (0.3209)	2.9182 (0.4626)	-0.1567 [11.6644]
Age	24.6072 (3.2752)	24.6989 (3.2207)	24.6641 (3.2880)	-0.2654 [2.6327]
Age Squared	616.2391 (164.3507)	605.6595 (160.6588)	619.1261 (165.3218)	-13.4666 [2.6618]
Percent Urbanization at Birth***	59.4919 (29.6548)	43.5173 (22.5608)	63.9919 (29.8618)	-20.4746 [23.3126]
Percent Urbanization at Birth Missing	0.0246	0.0	0.0313	-0.0313 [6.5861]

Continued

Table 2.2 - Variable Means Used in Migration Equation

Table 2.2 Continued

	Total	Appalachian	Non-Appalachians	Difference in Means [t-test]
Years in Original Labor Market Since Leaving School	5.6765 (3.2689)	6.3668 (3.4291)	5.4881 (3.1984)	0.8787 [8.7811]
Median Rent	392.6622 (117.3653)	322.4955 (64.6011)	411.8129 (121.1844)	-89.3174 [26.0092]

Source: National Longitudinal Surveys of Youth 1979 and 1990 Census, based on 1342 Appalachian person/years and 4918 non-Appalachian person/years. Standard deviations are in parentheses.
*Indicates variable is by county.
**Based on 1339 Appalachian observations and 4727 non-Appalachian observations.
***Based on 1342 Appalachian observations and 4764 non-Appalachian observations.
****Based on the 193 Appalachians and 986 non-Appalachians

	Total	Appalachian	Non-Appalachians	Difference in Means [t-test]
Employment Status	0.7971	0.7221	0.8151	-0.0930 [8.2789]
Migrant Status	0.1465	0.0876	0.1607	-0.0732 [7.3978]
AFQT**	-6.0891 (23.1445)	-17.2293 (21.3406)	-3.3160 (22.7405)	-13.9133 [21.9482]
AFQT Missing	0.0307	0.0044	0.0370	-0.0325 [6.7434]
Highest Grade Completed	11.9667 (2.6472	10.7392 (2.3891)	12.2622 (2.6214)	-1.5230 [21.0544]
Age	24.2411 (3.5191)	23.9563 (3.4538)	24.3010 (3.5314)	-0.3533 [35.844]
Age Squared	600.0150 (173.7817)	585.8258 (168.7375)	603.4230 (174.8126)	-17.5972 [3.6158]
Percent Urbanization at Birth***	59.7531 (29.6560)	43.5503 (22.7773)	63.7808 (29.7998)	-20.2306 [25.1837]
Percent Urbanization at Birth Missing	0.0253	0.0	0.0313	-0.0313 [7.1377]
Health Limitations	0.0393	0.0419	0.0387	0.0033 [0.5912]
Married Spouse Present	0.3734	0.3959	0.3679	0.0280 [2.0652]
Metro*	0.6658	0.5565	0.6921	-0.1357 [10.3152]
Unemployment Rate*	7.7169 (3.6670)	9.3875 (4.6367)	7.3146 (3.2676)	2.0729 [20.6680]
Lagged Unemployment Rate*	7.7483 (3.6144)	9.4075 (4.5099)	7.3488 (3.2391)	2.0587 [20.8034]

Continued

Table 2.3 - Variable Means Used in Employment Equation

Table 2.3 continued

Weekly Retail Earnings (in $100, 1990$)*	2.8064 (0.4678)	2.6172 (0.3658)	2.8525 (0.4781)	-0.2353 [18.2940]
Median Rent	397.7018 (117.9625)	327.6910 (75.1149)	414.5600 (120.1820)	-86.8690 [27.4328]

Source: National Longitudinal Surveys of Youth 1979, based on 1579 Appalachian person/years and 6574 non-Appalachian person/years. Standard deviations are in parentheses.
*Indicates variable is by county.
**Based on 1571 Appalachian person/years and 6331 non-Appalachian person/years.
***Based on 1579 Appalachian person/years and 6368 non-Appalachian person/years.

	Total	Appalachians	Non-Appalachians	Difference in Means [t-test]
Real Wage	870.7394 (792.9892)	729.7432 (611.2266)	902.1421 (824.8272)	-172.3989 [6.9557]
Migrant Status	0.1565	0.1014	0.1688	-0.0674 [5.9307]
AFQT**	-4.0556 (22.6291)	-14.6647 (21.2448)	-1.6133 (22.2320)	-13.0515 [18.7763]
AFQT Missing	0.0305	0.0040	0.0364	-0.0324 [6.0152]
Highest Grade Completed	12.1846 (2.6052	11.0161 (2.3570)	12.4449 (2.5866)	-1.4288 [17.8910]
Tenure	2.0014 (2.2850)	2.0616 (2.3868)	1.9880 (2.2617)	0.0735 [1.0258]
Tenure Squared	9.2260 (20.5004)	9.9421 (23.2938)	9.0665 (19.8236)	0.8756 [1.3619]
Experience	6.3929 (3.4079)	7.3781 (3.5924)	6.1734 (3.3326)	1.2047 [11.3764]
Experience Squared	52.4808 (52.5632)	67.3315 (60.3702)	49.1733 (50.0707)	18.1582 [11.1129]
Union	0.1376	0.1464	0.1356	0.0108 [0.9978]
Health Limitations	0.0318	0.0322	0.0317	0.0005 [0.0846]
Married Spouse Present	0.4025	0.4417	0.3938	0.0478 [3.1112]
Metro*	0.6682	0.5623	0.6918	-0.1295 [8.8152]
Unemployment Rate*	7.5955 (3.5940)	9.0777 (4.5284)	7.2654 (3.2606)	1.8123 [16.3902]

Continued

Table 2.4 - Variable Means Used in Wage Equation

Table 2.4 continued

Weekly Retail Earnings (in $100, 1990$)*	2.8068 (0.4678)	2.6172 (0.3658)	2.8525 (0.4781)	-0.2354 [18.2940]
Median Rent	400.7428 (118.7340)	333.1319 (78.5229)	415.8011 (120.9161)	-82.6692 [23.0456]
Lambda (w/ migrant status)	0.2495 (0.2411)	0.3418 (0.2734)	0.2290 (0.2282)	0.1128 [15.1722]
Lambda (w/ predicted migrant status)	0.2488 (0.2410)	0.3404 (0.2685)	0.2284 (0.2295)	0.1120 [15.0554]
Source: National Longitudinal Surveys of Youth 1979, based on 1202 Appalachian person/years and 5392 non-Appalachian person/years. Standard deviations are in parentheses. **Based on 1196 Appalachian person/years and 5193 non-Appalachian person/years.				

Constant	-3.2033 (0.9634)	
Appalachian native	-0.8180 (0.2525)	[-0.0078]
Afqt	0.0095 (0.0054)	[0.0001]
Afqt missing	0.0850 (0.3953)	[0.0065]
Highest grade completed	0.0834 (0.0378)	[0.0012]
Married spouse present	-0.5331 (0.1574)	[-0.0074]
Get married since last interview	0.7660 (0.1783)	[0.0275]
Get divorced since last interview	0.2961 (0.2411)	[0.0064]
Health limitations	-0.2876 (0.2885)	[-0.0031]
Metro*	0.1510 (0.1514)	[0.0021]
Lagged unemployment rate*	0.0856 (0.0336)	[0.0013]
Two period lagged unemployment rate*	-0.0117 (0.0420)	[-0.0002]
Weekly Retail Earnings (in $100, 1990$)*	-1.1122 (0.2420)	[-0.0166]
Born percent urban*	-0.0058 (0.0031)	[-0.0001]
Born percent urban missing*	-0.0189 (0.3928)	[-0.0003]

Continued

Table 2.5 - Migration Probit

Table 2.5 continued

Years in Original Labor Market Since Leaving School	0.0984 (0.0239)	[0.0015]
Median rent*	0.0060 (0.0007)	[0.0001]
Sigma	2.2584 (0.1946)	
Source: National Longitudinal Surveys of Youth 1979 and 1990 Census, based on 1342 Appalachian person/years and 4917 non-Appalachian person/years. Standard errors are in parentheses, marginal and delta effects are in square brackets. * Indicates variable is by county.		

Constant	-2.1379 (0.3391)		-2.1248 (0.3395)	
Appalachian Native	-0.0219 (0.1404)	[-0.0038]	-0.0563 (0.0958)	[-0.0099]
Appalachian*actual migration status	0.0877 (0.1373)	[0.0151]		
Non-Appalachian*actual migration status	0.0896 (0.0744)	[0.0154]		
Appalachian*predicted migration status			0.4634 (0.2490)	[0.0587]
Non-Appalachian*predicted migration status			0.0772 (0.0837)	[0.0128]
AFQT	0.0128 (0.0020)	[0.0022]	0.0127 (0.0020)	[0.0022]
AFQT Missing	0.2862 (0.1874)	[0.0411]	0.2845 (0.1880)	[0.0408]
Highest grade completed	0.2105 (0.0182)	[0.0362]	0.2127 (0.0181)	[0.0365]
Experience	0.1573 (0.0265)	[0.0270]	0.1700 (0.0230)	[0.0292]
Experience squared	-0.0035 (0.0016)	[-0.0006]	-0.0040 (0.0015)	[-0.0007]
Born percent urban*	-0.0046 (0.0013)	[-0.0008]	-0.0044 (0.0013)	[-0.0007]
Born percent urban missing*	-0.2044 (0.2312)	[-0.0397]	-0.1768 (0.2324)	[-0.0338]
Health Limitations	-0.1150 (0.1056)	[-0.0212]	-0.1154 (0.1057)	[-0.0212]
Married spouse present	0.2620 (0.0580)	[0.0432]	0.2635 (0.0580)	[0.0434]

Continued

Table 2.6 - Employment Probit

Table 2.6 continued

Metro*	-0.0216 (0.0737)	[-0.0037]	-0.0115 (0.0740)	[-0.0020]
Unemployment rate*	-0.0426 (0.0109)	[-0.0073]	-0.0412 (0.0109)	[-0.0071]
Lagged unemployment rate*	-0.0061 (0.0108)	[-0.0010]	-0.0048 (0.0108)	[-0.0008]
Weekly Retail Earnings (in $100, 1990$)*	0.2873 (0.0836)	[0.0494]	0.2752 (0.0833)	[0.0473]
Median Rent*	-0.0003 (0.0004)	[-0.0001]	-0.0004 (0.0004)	[-0.0001]
Sigma	0.8181 (0.0393)		0.8200 (0.0392)	
Source: National Longitudinal Survey of Youth, 1979. Based on 1576 Appalachian person/years and 6545 non-Appalachian person/years. Standard errors are in parentheses, marginal and delta effects are in square brackets * Indicates variable is by county.				

Constant	4.7320 (0.0993)	4.7796 (0.1008)
Appalachian native	-0.0782 (0.0421)	-0.0336 (0.0279)
Appalachian*actual migration status	0.0855 (0.0390)	
Non-Appalachian* actual migration status	0.0222 (0.0183)	
Appalachian*predicted migration status		0.0756 (0.0491)
Non-Appalachian* predicted migration status		0.0403 (0.0181)
Afqt	0.0053 (0.0006)	0.0051 (0.0006)
Afqt missing	0.0954 (0.0507)	0.0927 (0.0507)
Experience	0.0718 (0.0063)	0.0736 (0.0059)
Exprience squared	-0.0022 (0.0004)	-0.0023 (0.0003)
Tenure	0.0373 (0.0057)	0.0383 (0.0057)
Tenure squared	-0.0027 (0.0006)	-0.0027 (0.0006)
Highest grade completed	0.0735 (0.0054)	0.0718 (0.0054)
Union member	0.1613 (0.0154)	0.1607 (0.0154)
Health limitations	-0.0499 (0.0284)	-0.0501 (0.0284)
Married spouse present	0.0826 (0.0132)	0.0807 (0.0132)

Continued

Table 2.7 - Log Wage Equations

Table 2.7 continued

Metro*	-0.0280 (0.0181)	-0.0248 (0.0181)
Unemployment rate*	-0.0126 (0.0022)	-0.0115 (0.0021)
Weekly Retail Earnings (in $100, 1990$)*	0.1604 (0.0182)	0.1539 (0.0181)
Median Rent*	0.0003 (0.0001)	0.0003 (0.0001)
Lambda (w/actual migrant status)	0.3385 (0.0582)	
Lambda (w/predicted migrant status)		0.3122 (0.0573)
Rho	0.4019	0.4016
R-sq overall	0.3038	0.3037
Source: National Longitudinal Surveys of Youth 1979. Sample consists of 1243 Appalachian person-years and 5581 non-Appalachian person-years, with a total of 6824 person years. Standard errors in parentheses. * Indicates variable is by county.		

CHAPTER 3

A COMPARISON OF EMPLOYMENT, EARNINGS, AND MIGRATION OF ECONOMICALLY DISADVANTAGED WHITES USING MAXIMUM SIMULATED LIKELIHOOD

The empirical model is a wage equation with two additional endogenous regressors, migration and employment, estimated simultaneously using maximum simulated likelihood (MSL). The model is designed to investigate (i) how much of the Appalachian wage and employment gaps seen in the raw data can be attributed to differences in human capital, (ii) whether a wage penalty exists for being born and raised an Appalachian, regardless of current location, and (iii) whether migration is a plausible solution to poverty for Appalachians. Maximum likelihood estimates the parameters in the three equations simultaneously, ensuring that the estimates are consistent, efficient, and free of standard error bias that may result from multi-step estimation. Since the data are longitudinal, a random effect term is added to each equation to account for individual-specific unobserved variation. Correlation between the random effects is modeled, requiring a random effects covariance matrix to be estimated in addition to the model's other parameters. However, as the complexity of the model increases, so does the difficulty in both characterizing and ultimately solving the likelihood function. One tool available to facilitate solving detailed models is simulation. In this context, simulation replaces integrating each of the covariance matrix parameters over the state space at each iteration.

My results show a large part of the wage and employment differences are explained by differences in human capital, measured by years of schooling and AFQT scores. I find no evidence of an unobserved Appalachian trait that causes a wage penalty or a disassociation with the labor force. Thus, any effects on employment and wages caused by concentrated poverty or a culture of poverty are expressed through their impact on observed characteristics. In addition, migration does not appear to be a plausible solution for poverty, especially for Appalachians. There is evidence of self-selection into migration for this sample, where the most-able candidates for migration queue into labor market segments where the returns to skill are the largest. However, for a randomly selected Appalachian migration will not raise the wage to the level of a non-Appalachian EDW. This is supported by the low migration rates among Appalachians compared other EDWs raised elsewhere in the country. Since migration does not provide relief from poverty, the results suggest that the solution lies with human capital investments and development of economic infrastructure in the region. Since the data used in this chapter is the same as in chapter 2, the data section is omitted.

3.2 MSL method

The model is the same as in the last chapter. There are again three endogenous variables that correspond to the three equations in the model: a reduced-form discrete-time migration equation, a reduced-form employment equation, and a wage equation. There are again random effects term in each equation captures any individual-specific unobservable trait that may affect each dependent variable. However, including three random effects inside a simultaneous model requires two extensions that are not necessary in the multi-step method presented in the previous chapter. First, the likelihood is now inside a triple integral over the state space of each of the random variables. These integrals need to be evaluated at each iteration to produce the value of the likelihood function. Second, assuming the random effects are correlated, a random effects covariance matrix also needs to be estimated to account for correlation between the random effects. Correlation between the random effects could come in the form of individual-specific unobserved variables. For example, variations in innate ability could positively affect both wages and employment. Or, variations in a Jovanovic-style[23] job match term could positively affect employment and wage rates while negatively affecting migration. The second extension involves adding more parameters, while the first makes this model impossible to solve given the current level of technology. Simulation is used to replace the integral while estimating the likelihood without bias and with a variance that decreases as the number of simulations increases, via the law of large numbers. The combination of simulation and maximum likelihood is termed Maximum Simulated Likelihood (MSL).

MSL estimators are attributed to McFadden (1989), who suggested using simulation to solve the classical method of moments problem. McFadden noted that using the method of moments required solving an expected response function at each parameter value using numerical integrals, a process which becomes increasingly difficult as the number of discrete responses increases. For example, in a multinomial probit setting with m discrete responses, solving for the parameters requires the calculation of the $m - 1$ dimensional probabilities integrated over the parameter space. Stern (1994) offers this rationale for a generic simulation problem. Suppose simulation is used to evaluate

$$E[h(U)] = \int h(u) f(u) du ,$$

where U is a random variable with density $f(u)$, and some function $h(u)$ obtained from a economic model, such as the expected response functions mentioned above. Simulation takes R random i.i.d. draws from $f(u)$ and uses them to calculate a sample mean of $h(U)$. Then,

$$\frac{1}{R}\sum_{r=1}^{R} h(u^r)$$

is an unbiased estimator of $E[h(u)]$ with variance that approaches zero as the number of repetitions, R, approaches ∞.

The observation period for each individual begins the year that the individual first leaves school for at least 12 months. Each individual is followed until the year satisfying one of the following conditions: (i) the year after his first move, (ii) the year he first misses an interview, or (iii) the end of the sample frame in 1990 with continuous interviews. Since observations after the first move are omitted, return migration and second migration are not modeled in this paper.

At each period, a sample member's contribution to likelihood is determined by the combination of his observed employment and migration status. Thus, contributions to likelihood are split into four parts: migration and employed, migration and not employed, no migration and employed, and no migration and not employed. This chapter first describes the migration decision, then discusses the employment decision, and ends with the wages of employed workers.

3.2.1 The Migration Decision

Migration is a dichotomous choice determined by a latent migration variable in the spirit of Sjaastad (1962). The latent migration variable, m_{it}, is the present value of migrating to a new labor market net of remaining in the current labor market. The term m_{it} is specified as

$$m_{it} = X_{it}\beta + \varepsilon_{it}, \text{ where } \varepsilon_{it} = \eta_{it} + \nu_i .$$

X_{it} is vector of regressors, and the error term ε_{it} is decomposed into an individual-specific random effect term, η_i, and a white noise term, ν_{it}. Since m_{it} is not observed, M_{it} is defined as a dichotomous choice variable for individual *i,* indicating whether or not a migration out of his local labor market occurs at time *t*. M_{it} can be characterized as

$$M_{it} = 1 \text{ if } \eta_{it} > -X_{it}\beta - \nu_i$$

$$M_{it} = 0 \text{ if } \eta_{it} \le -X_{it}\beta - \nu_i .$$

The probability of migration in a single period becomes

$$Prob(M_{it} = 1 \mid X_{it}, \nu_i) = \Phi\left(\frac{X_{it}\beta + \nu_i}{\sigma_\nu}\right),$$

where Φ is the normal cumulative distribution function and σ_ν is the variance of the migration white noise term. Since the decision to migrate is made only once, the predicted probability that individual *i* has left the original labor market at any time before *t* is

$$Prob(M_{it} = 1 \mid X_{it}, \nu_i) = 1 - \prod_{\tau=1}^{t}\left[1 - \Phi\left(\frac{X_{i\tau}\beta + \nu_i}{\sigma_\nu}\right)\right].$$

The predicted probability that individual *i* has not left the original labor market at any time before *t* is

$$Prob(M_{it} = 0 \mid X_{it}, \nu_i) = \prod_{\tau=1}^{t}\left[1 - \Phi\left(\frac{X_{i\tau}\beta + \nu_i}{\sigma_\nu}\right)\right].$$

The contribution to the likelihood from migration is characterized as

$$Prob(\mathrm{M}_{\mathrm{i}\tau}) = \left\{\prod_{\tau=1}^{t}\left[\frac{1-\Phi(X_{it}\beta+\nu_i)]}{\sigma_\nu}\right]^{1-M_{i\tau}}\left[1-\prod_{\tau=1}^{t}\left[\frac{1-\Phi(X_{it}\beta+\nu_i)}{\sigma_\nu}\right]^{M_{i\tau}}\right]\right\} \tag{1}$$

3.2.2 The Employment Decision

The employment decision is determined in each period by the difference between two endogenous variables, the reservation wage and the wage offer (or market wage). Let w_{it}^{res} be the reservation wage for individual *i* in period *t*, where

$$w_{it}^{res} = \alpha_r Z_{it} + \delta_r M_{it} + \xi_{it},$$

$$\text{and } \xi_{it} = \omega_{it} + \kappa_i .$$

Z_{it} is a vector of regressors, M_{it} is the dichotomous migration variable, and the error term ξ_{it} is decomposed into an individual-specific random effect term, κ_i, and a white noise term, ω_{it}.

Similarly, let w_{it} be the wage offer,

$$w_{it} = \alpha_o Y_{it} + \delta_o M_{it} + \varsigma_{it},$$

$$\text{where } \varsigma_{it} = \vartheta_{it} + \varphi_i .$$

Y_{it} is a vector of regressors, and the error term ς_{it} is decomposed into an individual-specific random effect term, φ_i, and a white noise term, ϑ_{it}. If $w_{it}^R < w_{it}$, then the individual chooses to work. Let E_{it} represent the dichotomous employment choice variable. Using the equations for the wage offer and reservation wage, E_{it} can be characterized as $E_{it} = 1$ if $w_{it}^R < w_{it}$, or

$$\omega_{it} - \vartheta_{it} < (\alpha_o - \alpha_r)U_{it} + \alpha_o Y_{it} - \alpha_r Z_{it} + (\delta_o - \delta_r)M_{it} + \varphi_i - \kappa_i,$$

$$\text{and } E_{it} = 0 \text{ otherwise.}$$

To simplify the notation, I combine the regressors into $\overline{U}_{it}$ and random effects terms into ψ_i. The parameter vector, α, represents the difference between the structural parameters for variables in both Y_{it} and Z_{it}, the structural wage offer parameter for variables only in the wage offer equation, and the structural reservation wage parameter for variables only in the reservation wage equation. Symmetrically, δ is the difference between the structural wage offer migration return and the structural reservation wage migration return, so that

$$E_{it} = 1 \text{ if } \mu_{it} < \alpha \overline{U}_{it} + \delta M_{it} + \psi_i,$$

$$\text{and } E_{it} = 0 \text{ otherwise.}$$

Normalizing the variance of μ_{it} to one, the probability of no employment is

$$\text{Prob}(E_{it} = 0 \mid \overline{U}_{it}, M_{it}, \psi_i) = 1 - \Phi(\alpha \overline{U}_{it} + \delta M_{it} + \psi_i), \qquad (2)$$

which is the contribution to likelihood by those who are not working.

The contribution to likelihood from the employed is the product of the density (or marginal) of the wage offer error term, ς_{it}, and the probability of employment, conditional on the wage being observed. Specifically,

$$\text{Prob}(E_{it} = 1 \mid \ldots) = \left(\frac{1}{\sigma_\varsigma}\right)\phi(w_{it} - \alpha_o Y_{it} - \delta_o M_{it} - \varphi_i) *$$

$$\Phi\left(\frac{\alpha \overline{U}_{it} + \delta M_{it} + \psi_i - \dfrac{\rho}{\sigma_\varsigma}(w_{it} - \alpha_o Y_{it} - \delta_o M_{it} - \varphi_i)}{\sqrt{1-\rho^2}}\right) \qquad (3)$$

where ϕ is the standard normal probability distribution function, Φ is the standard normal cumulative distribution function, σ_ς is the variance of ς_{it}, and ρ is the covariance between ς_{it} and μ_{it}. The likelihood function is the product of the equations (1), (2), and (3) over all the observations.

3.3.3 Simulation and Random Effects

Simulation occurs at the random effects. To correct for time invariant unobserved individual effects such as family background, innate ability, or school quality, each of the three equations has an individual-specific random effects term labeled ν_i, ψ_i, and φ_i for the migration, employment, and wage equations, respectively. Without simulation, a triple-integral over the state-space of the random effects would be needed to include random effects in this model. These effects are comprised of three individual-specific draws from a trivariate normal, labeled u_{1i}, u_{2i}, and u_{3i}, and the Cholesky factorization of the covariance matrix, labeled *LL*'.

$$\begin{pmatrix} \nu_i \\ \psi_i \\ \varphi_i \end{pmatrix} = L \begin{pmatrix} u_{1i} \\ u_{2i} \\ u_{3i} \end{pmatrix}, \text{ where } L = \begin{pmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{pmatrix}.$$

Thus, $\nu_i = l_{11}u_{1i}$, $\phi_i = l_{21}u_{1i} + l_{22}u_{2i}$, and $\varpi_i = l_{31}u_{1i} + l_{32}u_{2i} + l_{33}u_{3i}$. The Cholesky factorization ensures that *L,* and thus the random effects covariance matrix *Ω = LL'*, is positive definite. Each of the three individual-specific random draws are taken one hundred times.

3.3.4 Identification

Both the migration and employment equations are reduced form, requiring variables specific to both but omitted from all other equations for identification. Thus, the reduced form migration equation includes all the variables in the employment equation, all the variables in the wage equation except job-specific variables, and a variable for percent county homeownership, which is excluded from the employment and wage equations. The reduced form employment equation includes all the variables in the migration equation, all the variables in the wage equation except job-specific variables, plus lagged county unemployment rate, which is excluded from the migration and wage equations. The wage equation is structural. The estimated parameters are l_{11}, l_{21}, l_{22}, l_{31}, l_{32}, l_{33}, ρ, σ_{ς}, σ_{v}, δ, and δ_{o}, and the vectors α_{o}, α, and β.

3.4 Results

Table 3.1 presents the estimates of the log wage equation. The second column presents estimates without random effects, estimated with Maximum Likelihood. The third column includes the random effect and is estimated by Maximum Simulated Likelihood. The dependent variable is the natural log of real hourly wages in 1990 dollars, so all wage equation coefficients represent a percentage marginal effect on real wages.

The Appalachian dummy in the wage equation is negative and insignificant, implying that the 22 percent difference in real wages is explained by differences in observed characteristics. Thus, there is no evidence of an unexplained wage penalty for Appalachians. Differences in human capital do explain a large part of the wage differential. Both highest grade completed and AFQT scores have positive and significant effects on wages. Under either specification, one additional year of schooling raises wages by 8 percent, while an additional point scored on the age-adjusted AFQT increases wages by just under 0.4 percent. The sample means show that Appalachians attain 1.45 less years of schooling and score 13 points lower on the AFQT. Therefore, the human capital proxies alone account for 53 percent of the wage differential using the average the Appalachian from the sample.[24] Experience also has a positive effect on wages. An additional year of experience increases wages by over 3 percent. Wages are sensitive to changes in the county unemployment rates, with an additional percent increase in the county unemployment rate lowering wages by just under 2 percent. Weekly county retail earnings are positive and significant, raising wages by over 24 percent for each hundred dollar increase in weekly retail earnings. This sensitivity of the retail earnings variable is a byproduct of drawing a sample of economically disadvantaged males who view the retail industry as a initial opportunity for employment. In addition, since wages are deflated using the national CPI, weekly retail earnings control for differences in prices at the local level.

Migration for both Appalachians and non-Appalachians produces no statistically significant wage returns. Since this sample is comprised of economically disadvantaged whites with below-average levels of human capital, these results support the idea of favorable self-selection among the more able and/or motivated as found in such studies as Borjas, Bronars & Trejo (1992) and Gabriel & Schmitz (1995). These authors find evidence that highly skilled workers are attracted to labor markets where the returns to skill are largest. Since this sample is comprised of economically disadvantaged whites with below-average levels of human capital, these workers presumably have fewer opportunities to increase wages through migration. Further, low returns to migration help explain why many Appalachians have not left the region. Because of low human capital levels, Appalachians as a group do not see migration as a solution to poverty, despite the region's close proximity to better economic environments in the South or Northeastern U.S.

This non-existence of a migration premium is in contrast to the previous chapter, which finds positive and significant returns (albeit small in absolute terms) to migrating for both Appalachians and non-Appalachians. These differences in returns to migration estimates are driven by the differences in estimation methods. Although both methods produce consistent estimates, the simultaneous estimation in the MSL method results in an increase of parameter efficiency since the standard errors are estimated consistently in one step.[25] Although all indications are that the instruments are not weak in the multi-step method, this is the risk involved in any instrumental variable approach. The predicted probability of migration in the multi-step method is particularly sensitive since it was used in two sequential steps: the employment equation and the wage equation.

Table 3.2 presents estimates of the employment equation. The Appalachian dummy in this equation is also negative and insignificant, implying that the 9 percent difference in employment is also explained by observed characteristics. The notion of an unobserved cultural disassociation to the labor force among the Appalachians is either absent or reflected in observed characteristics, in particular lower years of schooling and AFQT scores. Both highest grade completed and AFQT scores also have positive and significant effects on employment. One extra year of schooling raises the probability of employment by nearly 3 percent, while an additional point scored on the age-adjusted AFQT increases employment by 0.3 percent. County unemployment rates have a negative and significant effect.

Table 3.3 presents estimates of the migration equation. Again, the Appalachian dummy is negative and insignificant. The human capital proxies, AFQT score and highest grade completed, are both positive and significant. Thus, there is evidence of favorable self-selection among migrants in this sample. The county unemployment rate is negative and significant, suggesting that depressed local economic conditions increase the probability of migration by a slight amount.

Table 3.4 presents estimates of the Cholesky decomposition of the random effects covariance matrix. The estimates of the random effects error structure are all statistically insignificant. The off-diagonal terms are the closest to zero, suggesting no correlation between the random effects. Further, the parameter estimates are similar with and without random effects. A likelihood-ratio test between the two models shows no significant difference. These results imply that the model is over-specified for the amount of data available. Although not presented, the same model was run assuming no correlation between random effects via a diagonal covariance matrix. There was no significant difference in the two likelihoods.

3.5 Conclusions

This chapter and the previous use the NLSY79 to compare economic outcomes of economically disadvantaged white Appalachian males with those of economically disadvantaged white males born or raised elsewhere in the country. Both chapters develop an empirical model of migration, employment, and wages to assess the factors that explain employment and wage differences and to quantify the effects of migration on labor market outcomes of the two groups. This empirical model uses three endogenous variables: migration, employment, and wages. Using longitudinal data corrects for unobserved time-invariant effects correlated over time and across the three equations.

While the previous chapter used a multi-step approach to estimate the model, this chapter employs Maximum Simulated Likelihood (MSL). The main advantage of this approach over the multi-step method is simultaneous estimation, which produces consistent standard errors that do not correction at each step. Simulation is needed to include random effects in a three-equation model that is estimated simultaneously.

The raw data illustrates the Appalachian disadvantage in wages and employment. The employment rate of poor white Appalachians is 9.1 percentage points lower than other EDW men, and their wage is 22 percent lower. A large part of these differences is explained by differences in human capital alone, measured by years of schooling and AFQT scores. Measured regional differences in economic opportunity are also important determinants of the relatively poor economic performance of Appalachian men. Higher unemployment rates and lower weekly earnings in retailing inside Appalachia contribute substantially to Appalachian employment and wage deficits. There is no evidence that unmeasured characteristics associated with growing up in Appalachia adversely affect labor market outcomes. Thus, any effects on employment and wages of concentrated poverty or a culture of poverty are expressed through their impact on observed characteristics such as years of schooling and AFQT scores.

As a whole, migration is not an answer to poverty for randomly selected Appalachians, despite the region's close proximity to economically viable areas. There is evidence that favorable self-selection in migration exists for those with high levels of human capital. Since the sample has below-average levels of human capital, the expected wages from migrating to a new labor market do not appear to be consummately higher.

This is particularly true for Appalachians, who have lower levels of human capital even when compared with economically disadvantaged whites raised elsewhere in the country. This result is in contrast to the multi-step estimation migration premium, which finds a small absolute gain to migrating. This is likely the result of inefficiency in the multi-step approach exacerbated by the inclusion of the migration instrumental variable in two sequential steps.

	No random effects	Random effects
Constant	4.7076 (0.1520)	4.6987 (0.1521)
Appalachian native	0.0135 (0.0422)	-0.0012 (0.0418)
AFQT	0.0033 (0.0009)	0.0044 (0.0011)
Highest grade completed	0.0824 (0.0074)	0.0794 (0.0074)
County weekly retail earnings (in $100, 1990$)	0.2452 (0.0310)	0.2505 (0.0316)
County unemployment rate	-0.0164 (0.0045)	-0.0177 (0.0045)
Experience	0.0358 (0.0065)	0.0342 (0.0071)
Appalachian migration dummy	-0.1410 (0.1263)	-0.1311 (0.1268)
Non-Appalachian dummy	0.0009 (0.0485)	0.0035 (0.0496)
Error variance	0.4832 (0.0099)	0.4806 (0.0103)
Source: National Longitudinal Surveys of Youth 1979 cohort and 1990 Census, based on 1342 Appalachian person/years and 4917 non-Appalachian person/years. Standard errors are in parentheses.		

Table 3.1 - MSL Estimates, Wage Equation

	Random effects		No random effects	
Constant	-0.3571 (0.3551)		-0.3667 (0.3853)	
Appalachian dummy	0.0396 (0.1018)	[-0.0007]	0.0446 (0.1026)	[0.0137]
AFQT	0.0108 (0.0023)	[0.0033]	0.0103 (0.0022)	[0.0032]
Highest grade completed	0.1095 (0.0195)	[0.0282]	0.1000 (0.0203)	[0.0275]
Appalachian migration dummy	0.0619 (0.3926)	[0.0185]	0.0672 (0.4094)	[0.0189]
Non-Appalachian dummy	-0.1819 (0.2048)	[-0.0545]	-0.1791 (0.2137)	[-0.0540]
County weekly retail earnings (in $100, 1990$)	0.1053 (0.0868)	[0.0250]	0.1018 (0.0867)	[0.0243]
County unemployment rate	-0.0896 (0.0191)	[-0.0234]	-0.0887 (0.0191)	[-0.0229]
Lagged county unemployment rate	0.0423 (0.0189)	[0.0090]	0.0460 (0.0188)	[0.0105]
Rho	-0.5759 (0.0520)		-0.5762 (0.0522)	

Source: National Longitudinal Surveys of Youth 1979 cohort and 1990 Census, based on 1342 Appalachian person/years and 4917 non-Appalachian person/years. Employment error variance normalized to one. Rho is the correlation between the employment and wage equation white noise terms. Standard errors are in parentheses, marginal effects are in square brackets.

Table 3.2 - MSL Estimates, Employment Equation

	Random effects		No random effects	
Constant	-0.6561 (0.7051)		-0.7121 (0.6863)	
Appalachian dummy	-0.0590 (0.1454)	[-0.00006]	-0.0699 (0.1415)	[-0.00006]
AFQT	0.0077 (0.0030)	[0.000007]	0.0071 (0.0029)	[0.000007]
Highest grade completed	0.0535 (0.0241)	[0.00006]	0.0549 (0.0235)	[0.00005]
County weekly retail earnings (in $100, 1990$)	-0.4133 (0.1204)	[-0.0004]	-0.4014 (0.1172)	[-0.0004]
County unemployment rate	-0.0561 (0.0163)	[-0.00007]	-0.0530 (0.0156)	[-0.00006]
County home-ownership rate	-0.0263 (0.0053)	[-0.00002]	-0.0257 (0.0051)	[-0.00002]
Error variance	1.4717 (0.3101)		1.4801 (0.3231)	
Source: National Longitudinal Surveys of Youth 1979 cohort and 1990 Census, based on 1342 Appalachian person/years and 4917 non-Appalachian person/years. Standard errors are in parentheses, marginal effects are in square brackets				

Table 3.3 - MSL Estimates, Migration Equation

	Random effects	No random effects
L11	0.203 (0.282)	
L21	-0.0913 (0.128)	
L22	-0.0585 (0.101)	
L31	-0.0443 (0.0411)	
L32	0.0024 (0.0306)	
L33	-0.0066 (0.0347)	
Log likelihood	4577.6962	4576.127

Note: Standard errors are in parenthesis.

Table 3.4 - MSL Estimates, Random effects structure

CHAPTER 4

THE APPALACHIAN WAGE GAP, 1940-1990

Up to this point, the empirical focus has a comparison of economically disadvantaged white males raised in Appalachians and the rest of the country using data from 1979 to 1990. Using IPUMS (Integrated Public Use Microdata Series) Census data from 1940 to 1990, this chapter broadens the analysis to include all full-time members of the workforce in Appalachia. I also construct a control group for this chapter using all full-time workers living elsewhere in the country. The sample frame is also significantly larger, which will shed some light on the general wage patterns inside and outside of the region since 1940. Data from the IPUMS Census project show Appalachian log weekly wages were 18% lower than the non-Appalachian sample in 1940, and by 1990 the wage gap grew to 22% (See Table 4.1). Between 1940 and 1970 the wage gap stayed relatively constant as wages grew considerately both inside Appalachia and in the rest of the country. Stagnant wage growth during the 1970s outside of the region helped Appalachians close the wage to 13.5% by 1980. However, the 1980s erased all of the Appalachian gains in the previous decade as wages inside the region did not grow while non-Appalachian weekly wages grew by about 8%.

Educational deficits also appear in the IPUMS data. Appalachians have roughly 0.8 years less of schooling than the rest of the country (See Table 4.2). Clearly the wage and education deficits between Appalachia and the rest of the country are related. It is possible that Appalachians have less educational attainment because of lower returns to education within the region. A natural way to investigate this relationship is using Mincerian-style wage regressions on national data. Table 4.3 presents some of the results from regressions of log weekly wages on an Appalachian dummy, education, experience, black, female, industry and occupation controls decennially from 1940 to 1990. Table 4.3 also presents estimates from the same model with the Appalachian dummy and education interacted. The results show that the observable characteristics of the regressions do a poor job explaining the Appalachian wage gap. The Appalachian dummy in all of the regressions except one are roughly the same size as the observed log weekly wage gap. Returns to schooling do not appear to be a driving force in the wage gap from 1940 to 1980. In fact, Appalachians receive slightly higher returns to education on the margin from 1940 to 1980. These results are similar to the Mexican-white wage gap analysis in Trejo (1997), who finds similar returns to schooling for Mexicans and whites despite a large gap in wages between the two groups.

The exception is the interacted Appalachian and schooling regression from 1990. The regression without the Appalachian dummy/schooling interaction term looks similar to those in previous years. The Appalachian dummy is negative, statistically insignificant, and near the observed wage gap for 1990. By including the interaction term, the Appalachian dummy becomes statistically insignificant. In addition, non-Appalachians have higher marginal returns to schooling for the first time in the regressions. Thus, differences in returns to schooling fully explain the wage gap in 1990. A likely suspect for this marked change in the estimates is the large increase in the returns to skill during the 1980s found in several studies (Juhn, Murphy & Pierce (1993), Bound & Johnson (1992), and Murphy & Welch (1992) to name just a few) coupled with the skill gap between Appalachia and the rest of the country. Taken together, these two facts suggest that the rise in the returns to skill during the 1980s has been largely unfavorable to Appalachians.

This chapter investigates the driving forces of the wage gap between 1940 and 1990. Using a model formulated by Oaxaca (1973) and Blinder (1973), I decompose changes in the wage gap between the Appalachia and the rest of the country into changes in quantities of observable characteristics, changes in the skill prices associated with the observables, and changes in the wage equation residual for each ten-year period between 1940 and 1990. Skill prices are measured using Mincerian-style wage regressions for the control group (non-Appalachians) in each year. The observable characteristics used are education, labor market experience, black, female, industry and occupation. Originally used for black-white and male-female wage gaps, the Oaxaca and Blinder decomposition model illustrates how changes in both the skill differential and skill prices can alter the wage gap. This model appears well-suited for a sample of Appalachians and non-Appalachians since the previous chapters find a skill differential between the two groups and skill prices have fluctuated substantially between 1940 and 1990, particularly after 1970 (see Juhn, Murphy, & Pierce (1993)).

In addition to including a standard set of human capital proxies (education and experience) and demographic controls (black and female), I also include broad industry and occupation dummy variables in the set of observable characteristics. These variables are included because of the differences in industry and occupation makeup between Appalachia and the rest of the country. In the early part of the twentieth century the regional economy of Appalachia relied heavily on extraction of natural resources, particularly coal. Although coal remains an important resource, it is not a major provider of jobs. Mining and agriculture accounted for only 5.1% of total Appalachian worker hours for the region in 1990.[26] Feser & Goldstein (2002) detail the Appalachian lag in high technology firms, arguing that a lower supply of high paying jobs exists in the region.

Several studies point out the impact of industry and occupation on wages, including Krueger & Summers (1987 and 1998), Dickens & Katz (1987), and Gibbons & Katz (1992). Kaboski (2002) shows that much of the wage growth experienced over the past fifty years is a combination of higher human capital levels *and* workers switching into higher paying jobs created by large increases in technology. Several causes of differences in wages across industries have been proposed: compensating differentials between industries (Viscusi & Moore (1991), Hersch (1998)), unmeasured ability across industries (Murphy & Topel (1987)), the presence of unions (Card (1996)), or higher returns to skill in high tech jobs (Juhn, Murphy, & Pierce (1993) and Murphy & Welch (1993)). I do not discuss the cause of inter-industry and inter-occupation wage differentials in this paper, rather acknowledging their impact on wage regressions that drive the Oaxaca and Blinder decomposition model.

Juhn, et al. (1991) extend the original Oaxaca and Blinder model by decomposing the wage residual in two different ways depending on its interpretation. In one setting, the wage residual decomposition evaluates the effect of income inequality on the wage gap. If the wage residual is interpreted as payments to unobservable skill and assuming the distribution of unobservable skill is constant, then the wage gap of effect changes in the return to unobservable skill are quantified. The other residual decomposition uses the assumption that differences in skill between Appalachia and the rest of the country are caused solely by differences in school quality.

It is possible that the Appalachian wage gap is driven by its primarily rural makeup. If this is the case, then the Appalachian wage gap is simply an example of the wage gap that exists between urban and rural areas. To test for this possibility, an additional control group of rural non-Appalachians was drawn. Mean weekly wages for rural non-Appalachians in 1990 are less than two percent higher than Appalachians, suggesting much of the wage differential between Appalachia and the rest of the country can be attributed to differences in urban-rural composition. However, Appalachians do appear to have less skills than other rural workers in the U.S. Mean education levels for rural non-Appalachians are roughly 0.5 years higher than Appalachians throughout the sample frame. Although the wage gap is virtually non-existent in 1990, rural non-Appalachians did have had higher wages than Appalachians in every other sample year except for 1940. The largest wage gap between these groups was 1970 was just under nine percent in 1970. Since 1970, rural non-Appalachians wages have fallen slightly while Appalachian wages increased, leaving the wage gap at just under two percent in 1990. Since there are fluctuations in the wage gap between these groups, the decompositions done a second time using the rural non-Appalachian control group.

I use the same data set to estimate changes in labor demand for each sample using the same method as Murphy & Katz (1992). The model uses national trends in industries and occupations and weights them by regional shares to produce a change in labor demand index for each sample.

4.2 Data

The data are taken from the Integrated Public Use Microdata Series (IPUMS) Census project. The data points are answers based on the year prior to the Census. IPUMS data begins in 1850, but did not begin to collect schooling attainment and wage information until 1940. Thus, the sample range for this paper is decennial 1940 to 1990. At this time, the 2000 data is only available in "beta" form and thus still under testing for consistency. Several sample densities are available, but only the 1% density used in this paper is available in each sample year 1940 through 1990. The variables of interest are wages, education, experience, occupation, industry, race, gender, and location.

Weekly wages are constructed by dividing the IPUMS variable annual wage and salary income by the number of weeks worked. Other IPUMS income variables are available but were not used since I am interested only in wages from employment and not rents from capital or transfer payments. In 1960 and 1970 only ranges for weeks worked last year were collected, called "intervalled weeks worked" in the data. Intervalled weeks worked data are coded into 6 ranges between zero and 52. The midpoint of the range is used to determine weekly wages. Observations with less than 40 weeks worked in the previous year or less than 35 hours worked last week (as identified by hours worked last week IPUMS variable) were omitted to limit the sample to full-time workers. IPUMS top codes salary wages each year by assigning the mean wage for all respondents above a given wage. I eliminated observations with wages greater than 500 dollars per hour or less than half of the sample year's federal minimum wage. Thus, everyone in the sample was drawing a wage at some point during the year prior to the Census. The 1990 GDP deflator was used to construct real wages across sample years.

Education levels come from the educational recode IPUMS variable. Highest grade completed and attended were collected from 1940 to 1980, but in 1990 the Census used coded values for education levels less than 8th grade and began asking for highest degree attained for those with at least a high school diploma. Educational recode was created to be a consistent measure across 1940 to 1990. The ranges for the educational recode variable are less than first grade, between first and fourth grades, between fourth and eighth grades, ninth grade, tenth grade, eleventh grade, twelfth grade, between one and three years of college, and four years and higher of college.

After schooling level is calculated, experience is calculated using age minus highest grade completed minus 6. Experience levels are grouped into ten year intervals up to sixty. The highest experience level classification is sixty years and higher.

The classification system for occupation utilizes nine broad classes using the IPUMS variable occupational code, 1950 basis. These categories are listed in Table 4.3.

Similarly, the classification system for industry utilizes sixteen broad classes using the IPUMS variable industry code, 1950 basis. These categories are listed in Table 4.6. Since industry and occupation codes change over time, IPUMS created the 1950 basis for each variable to ease comparison across the years. Ideally, the occupation codes describe the type of work done by the worker and the industry codes detail the sector of employment. In both the occupation and industry variables, a higher level of detail exists in the data. However, since the data are based on the descriptions of the respondent and interpretation by the coder, a significant amount of measurement error exists. Using the broad occupation and industry system removes some of this measurement error while preserving the differences in mean education and wages across broad industries and occupations.

The geographic variables in the IPUMS have varied over the years. In general, the data do not identify a geographic area with less than 100,000 residents. Specifically, in 1940 and 1950 the smallest geographic unit was city of residence, provided the city had more than 100,000 residents. If a respondent did not live in a city with at least 100,000 people, the smallest identifiable unit is the state. In 1960, the smallest identifiable geographic units are states. From 1970 to 1990, county groups (called Public Use Microdata Areas, or PUMAs in 1990) are available, which are constructed to have no less than 100,000 residents. Despite this variation in geographic variables, a metro area variable exists in each sample year that identifies respondents living inside the central city, outside the central city, or in a rural area.

Since Appalachia is defined by county and is spread over thirteen states, the variation in geographic variables prevents a precise definition of the region. Thus, the Appalachian sample in this paper relies on IPUMS state and metro area variables. In addition to respondents in West Virginia (the only state wholly within Appalachia), I include respondents in the rural areas of Ohio, Pennsylvania, Kentucky, Tennessee, and Alabama in the Appalachian sample. Each of these states contains a large rural portion inside Appalachia. The other states that are partially contained in Appalachia (New York, Maryland, Virginia, North Carolina, South Carolina, Georgia, and Mississippi) are excluded because of the relatively small rural area within Appalachia. This definition of the region is predominantly rural since it omits some of the largest of Appalachia's urban areas; Pittsburgh, PA, Birmingham, AL, Knoxville, TN, and Chattanooga, TN. Since there are urban areas both within and outside of Appalachia in each of these states, respondents living in these cities cannot be identified. All urban areas in the Appalachian sample are in West Virginia, such as Charleston, Huntington, and Wheeling. Unfortunately this definition most likely includes a small number of people living outside of the region by including rural parts of the aforementioned states. Because the states chosen have small rural portions outside of Appalachia and sampling in rural areas is particularly sparse in the 1% density samples, this is not thought to be problematic.

The non-Appalachian sample includes respondents living in all of the states with no portion in the Appalachian region, plus those living in urban parts of Ohio, New York, Georgia, Maryland, Mississippi, Virginia, and Kentucky.[27] The 1960 data does not identify these cities explicitly, but they can be indirectly determined using the IPUMS metropolitan area variable that lists urban areas above a given population level.[28] Although this variable does not exist for 1960, no urban areas in these states were in the Appalachian region in any other sample year.[29] Thus, this definition of the non-Appalachian region captures the urban areas that border Appalachia in addition to those parts of the country not adjacent to the region. North Carolina and South Carolina were not included in either sample because each contains urban areas inside and outside of Appalachia, which prevents identifying Appalachian status with only state and metro area data.

The metropolitan variable mean shows that 91% of the Appalachian sample lives in a rural area while the non-Appalachian sample has roughly 30% living in rural areas over the entire sample frame.[30] To test whether the Appalachian wage and human capital deficits are byproducts of the tendency for those in rural areas to have lower wages and skills, a sample of rural non-Appalachians was created. This sample is comprised of all respondents in the full Appalachian sample living in an IPUMS-defined rural area.

Table 4.1 presents the sample means of real log wages by year. Appalachians earn less than non-Appalachians in each of the sample years, a gap that has increased over the sample frame. Both groups experienced the same high levels of real wage growth from 1940 to 1970, and each stalled to near zero growth in 1980. During the 1980s the non-Appalachians sample experienced large wage growth while Appalachian wage growth remained stagnant. By the end of the 1980s Appalachian wages were roughly 22% lower than non-Appalachian wages, producing the largest gap in the sample frame. From a wage standpoint alone, Appalachians are not particularly disadvantaged after accounting for their predominantly rural status. Appalachians and rural non-Appalachians had nearly identical wages in 1940 and 1990. All three samples experienced a sharp decrease in wage inequality during the 1940s and a sharp increase during the 1980s, leaving the income inequality in 1990 higher than the 1940 levels for each sample.

The mean education levels in Table 4.2 are similar to the wage means. Education levels for all groups rose steadily across 1940 to 1990, each gaining about three years during the sample frame. The ordering of education means stays constant throughout the sample frame; non-Appalachian education means are the highest, rural non-Appalachians are second, and Appalachian education means are the lowest. The absolute gap in education stays roughly constant throughout the sample frame. In 1940, the mean Appalachian education level is 0.6 years less than non-Appalachians and 0.8 years less in 1990.

Table 4.2 also presents the means of labor market experience, black, and female. The non-Appalachian sample has higher mean percentages of blacks and females in the workforce throughout the sample frame. Female participation grew considerably over the sample frame, roughly doubling the 1940 rates by 1990. There is very little variation in experience between the samples, particularly between 1940 and 1970. Appalachians have slightly higher experience levels than non-Appalachians in 1980 and 1990, amounting to roughly a 0.1 differential, or about one year since experience is measured in ten-year intervals.

Table 4.4 presents mean wage levels in broad occupation groups for the combined Appalachian and non-Appalachian samples, delineated by year. The broad occupations with the highest mean wages and education levels throughout the sample frame are professional and technical and managers, officials, and proprietors. These two broad occupation categories are predominantly white collar, including lawyers, accountants, and engineers. The lowest paying broad occupation groups throughout the sample are farmers and service workers, paying at least one dollar less in real wages than the next highest paying occupation for every sample year.

Table 4.5 presents the participation shares in each broad occupation by year. In general, the highest paid occupations experience increasing participation rates over time. Managers, officials, and proprietors, the highest paying occupation over the sample frame, grows from 8.1% to 14.9% over the same time frame. The same trend works in the opposite direction for low paying jobs. Operatives, a predominantly blue-collar occupation that includes bus drivers, janitors, and waiters/waitresses, has by far the largest non-Appalachian participation rate in 1940 of 18.9%. By 1990, participation falls almost in half to 11.8%. Although Appalachians exhibit increasing participation rates in the higher paid occupations, the occupation shares in the highest paid occupations are considerately lower. In 1990, Appalachian participation in the two highest paid occupations (managers, officials, and proprietors and professional and technical) is ten percentage points lower than non-Appalachian participation rates. Rural non-Appalachian participation rates in the highest paying occupations fall between the Appalachian and non-Appalachian participation rates.

Tables 4.6 and 4.7 present a similar analysis for broad industry groups. The broad industry means are not as divergent as the means for broad occupation. However, the same trends seen in the occupation means exist for industry means; higher paying industries have a higher percentage of non-Appalachians and workers in all three samples have increasing participation rates in the highest paid industries. Further, a significant amount of variation in industrial composition exists between each sample. The two manufacturing industries, durable and non-durable, employ 34 percent of the Appalachian worker hours in 1990 compared with 20.7 percent of the non-Appalachian and 23.2 percent of the rural non-Appalachian worker hours.

4.3 Decomposition Methodology

The methodology originates from Oaxaca (1973) and Blinder (1973) who investigate black-white and male-female wage gaps in the 1970s and 1980s. More recent studies using this decomposition are Blau & Kahn (1997) (male-female differentials) and Juhn, et al. (1991) (black-white differentials). Using estimates from Micerian-style wage regressions, this method decomposes wage convergence into contributions from changes in observable characteristics, changes in skill prices, and changes in the residual. The wage equation residual is interpreted as payments to unobservable factors, such as ability, family background, or school quality. In addition to evaluating the effect from changes in observable characteristics and skill prices, Juhn, et al. (1991) and Blau & Kahn use this decomposition to study discrimination in labor markets. Since discrimination is not thought to be a driving force in Appalachian/non-Appalachian wage gap, it is assumed that it plays no role in wage determination. Although the terminology below will focus on the non-Appalachian control group, I use two control groups, rural non-Appalachians, and a full sample of non-Appalachians.

This method begins with a wage equation for the control group of non-Appalachians.

$$Y_{it}^{NA} = X_{it}^{NA}\beta_t^{NA} + \sigma_t\theta_{it}^{NA},$$

where X_{it} is a vector of observable characteristics including highest grade completed, experience in ten-year intervals, black, female, and broad industry and broad occupation dummy variables. The wage residual specification is the contribution of Juhn, et al. (1991) to the original model. θ_{it} is a random variable with mean zero and variance one for all *t*, and σ_t is the non-Appalachian standard deviation of wages. The standard deviation of wages will serve as a measure for residual wage inequality later in the model. This method does not impose normality on the error structure. The subscript *t* indexes the six decennial Censuses between 1940 and 1990.

Analogously, a wage equation for Appalachians containing the same set of observables using the non-Appalachian skill prices is written. The wage gap in year *t*, defined as D_t, is

$$\begin{aligned} D_t = Y_{it}^{NA} - Y_{it}^{A} &= X_{it}^{NA}\beta_t^{NA} + \sigma_t\theta_{it}^{NA} - (X_{it}^{A}\beta_t^{NA} + \sigma_t\theta_{it}^{A}) \\ &= (X_{it}^{NA} - X_{it}^{A})\beta_t^{NA} - \sigma_t(\theta_{it}^{NA} - \theta_{it}^{A}) \end{aligned}$$

In year *t*, the wage gap between Appalachians and non-Appalachians can be decomposed into the difference in observable characteristics weighted by the non-Appalachian skill prices and the residual gap.[31] To evaluate wage convergence between two years, consider the future date of *t'*. Wage convergence is the difference in the wage gap between year *t* and year *t'*.

$$D_t^{'} - D_t = [(X_{it'}^{NA} - X_{it'}^{A}) - (X_{it}^{NA} - X_{it}^{A})]\beta_{t'} + (X_{it}^{NA} - X_{it}^{A})(\beta_{t'} - \beta_t) +$$

$$[(\theta_{it'}^{NA} - \theta_{it'}^{A}) - (\theta_{it}^{NA} - \theta_{it}^{A})]\sigma_{t'} + (\theta_{it}^{NA} - \theta_{it}^{A})(\sigma_{t'} - \sigma_t) \quad (4.1)$$

The first term is the contribution to wage convergence from changes in the gap of observable characteristics, holding constant the prices (i.e., wage regression coefficients) of observable characteristics.[32] The second term captures the effect of changes in non-Appalachian prices of observable characteristics, holding constant the gap in observable characteristics. Juhn et al. (1991) refers to the third term as the gap effect, which measures the effect of changes in the relative wage positions of Appalachians and non-Appalachians holding constant non-Appalachian residual inequality, changes in observable characteristics, and changes in prices. The unobserved prices effect, captured by the fourth term in equation (4.1), measures the effect of changes in non-Appalachian unobservable prices on wage convergence, holding the percentile ranking of Appalachians on the non-Appalachian residual distribution constant.

The focus of the results is on the first, second, and fourth terms. The first term is termed the quantity effect of observable characteristics since it measures the impact of changes in differences of observable characteristics on the wage gap. For example, this term will quantify the impact of higher non-Appalachian participation rates in the highest paying industries and occupations seen in the sample means. The second term, named the price effect, examines how changes in the coefficients of the wage regressions affected the wage gap. Among other things, this term will evaluate the importance of increasing returns to high levels of education during the 1980s to the wage gap. The wage regressions discussed in the beginning of this chapter suggest the price effect during the 1980s should be nontrivial.

An attractive feature of the Juhn, et al. (1991) extension to the original decomposition model is the identification of income inequality effects. Juhn, et al. (1991) named this term the unobserved prices effect since it captures the effect of changes in the residual distribution, i.e., changes in the payments to various forms of unobservable characteristics. An increase in income inequality in the non-Appalachian wage distribution will increase the wage gap even if Appalachians hold the same percentile ranking in the wage distribution. For example, the disproportionately high returns to high skilled workers during the 1980s caused both the prices of skill and income inequality to increase substantially. Appalachians that held the same percentile ranking in the non-Appalachian wage distribution in 1980 and 1990 would experience a rising wage gap by virtue of being on the left side of the wage distribution, ceteris paribus. The opposite is also true during periods wage inequality decreased such as the 1940s. By construction of the model, the effect of changes in inequality net of the contributions from observable characteristics is identified.

Interpretation of the residual components relies on several strong assumptions. Using the standard wage equation framework, the residual represents the prices of all unobservable components that affect wages. This includes the usual set of unobservables that potentially affect wages: ability, family background, school quality, job match quality, etc. Blau & Kahn point out that this does not take into account changes in the sample composition, equation misspecification, measurement error, or any productivity characteristics that affect non-Appalachian wages. For example, if the wage residual is interpreted as payments to unobservable skill (e.g., ability), then the results rely on the assumption that the distribution of unobservable skill is constant over time. Changes in sample composition may be problematic for the IPUMS data used here, since a different sample is drawn in each Census year. On the other hand, the large sample size available from in the IPUMS data diminishes the effects of changes in sample composition.

Juhn, et al. (1991) provide additional focus on school quality. Although the exact relationship between school quality and earnings later in life is under debate, there is a consensus in the literature of substantial variation in school quality. Card & Krueger (1992) find variation across states using common proxies for school quality (student/teacher ratio, length of school year, and teacher wage). State school quality proxies presented in their tables are show that most Appalachians states are far below the national average. In order to quantify this effect, I follow Juhn, et al. (1991) and begin with the assumption that the skill differential between Appalachia and the rest of the country is entirely captured by differences in school quality. The first step is to isolate education in the non-Appalachian wage equation.

$$Y_{it}^{NA} = S_{it}^{NA}\delta_t^{NA} + Z_{it}^{NA}\alpha_t^{NA} + u_{it}E(u_{it} \mid S_{it}^{NA})$$

δ_t^{NA} is the non-Appalachian return to schooling and Z_{it}^{NA} is a vector of all remaining observable characteristics. Let $S_t^{A^*}$ be the mean "quality-adjusted" education level for Appalachians measured in non-Appalachian education years, such that $S_t^{A^*} = S_t^A - Q_t$, where Q_t is the gap between observed schooling and "effective" schooling for Appalachians. Using $S_t^{A^*}$ in the wage gap decomposition, equation (4.1) becomes

$$D_t^{'} - D_t = [(Z_{it'}^{NA} - Z_{it'}^{A}) - (Z_{it}^{NA} - Z_{it}^{A})]\alpha_{t'} + [(S_{it'}^{NA} - S_{it'}^{A^*}) - (S_{it}^{NA} - S_{it}^{A^*})]\delta_{t'} +$$

$$(Z_{it}^{NA} - Z_{it}^{A})(\alpha_{t'}^{NA} - \alpha_{t}^{NA}) + (S_{it}^{NA} - S_{it}^{A^*})(\delta_{t'}^{NA} - \delta_{t}^{NA}) \,. \qquad (4.2)$$

The first term (quantity effects) and third terms (price effects) carry the same interpretation as in equation (4.1), except that schooling is not part of Z_{it}^{NA}. The second term is the effect of the changes in the gap between non-Appalachian schooling and effective Appalachian schooling, and the fourth term is the effect of changes in the prices of schooling. Substituting for $S_t^{A^*} = S_t^A - Q_t$ in equation (4.2), the decomposition becomes

$$D_t^{'} - D_t = [(X_{it'}^{NA} - X_{it'}^{A}) - (X_{it}^{NA} - X_{it}^{A})]\beta_{t'} + (X_{it}^{NA} - X_{it}^{A})(\beta_{t'}^{NA} - \beta_{t}^{NA}) +$$

$$(Q_{t'} - Q_t)\delta_{t'}^{NA} + Q_t(\delta_{t'}^{NA} - \delta_{t}^{NA}) \qquad (4.3)$$

The first and second terms are the same as in equation (4.1). The third term is the effect of changes in the Appalachian schooling gap between observed and effective schooling. The fourth term is the price effect of differences between observed and effective schooling. The school quality results will focus on the fourth term. Under the school quality interpretation of the residual, the fourth term summed together with the changes in school prices term to produce a quality-adjusted schooling prices effect on the wage gap.

This extension requires solving for $S_t^{A^*}$ and then Q_t. Equation (4.4) illustrates this process.

$$S_t^{A^*} = \frac{1}{N}\sum_{i=1}^{N}\frac{w_{it}^{A} - Z_{it}^{A}\alpha_t^{NA}}{\delta_t^{NA}} \tag{4.4}$$

In practice, each Appalachian's observable characteristics (except for schooling) are inserted into in the non-Appalachian wage equation. $S_{it}^{A^*}$ is the solution of this equation at the individual-level. The mean over all Appalachians in a given year (*N* in equation 4.4) produces $S_t^{A^*}$. In essence, the process determines the mean amount of schooling a non-Appalachian would have given Appalachian wages and observable characteristics.

4.4 Decomposition Results

The results for the decomposition between the non-Appalachian and Appalachians samples are presented at Tables 4.8. The rural non-Appalachian and Appalachian decomposition is presented at Table 4.9. The first row in each table is the wage convergence during the time period. This term is the difference in the log wage gap over the two sample years. By construction of the model, a positive number in Tables 4.8 and 4.9 represents changes unfavorable to Appalachians; i.e. contributions that increase the wage gap. The "all observables" row gives contributions to wage convergence from changes in prices and quantities of observables and represents the first and second terms from equation (4.1). This row is further decomposed into changes in observables only (first term, eqn. (4.1)) and prices only (second term, eqn. (4.1)). The "residual" row is the difference between wage convergence and contributions from all observables. The unobserved prices effect is the fourth term from eqn. (4.1), and the school quality effect is the fourth term from equation (4.3). Unlike the quantity and price effects, these two terms are constructed using different interpretations of the residual and should not be considered together.

Summing up the log wage differentials over all the sample years in Table 4.8 shows that the non-Appalachian/Appalachian wage gap has increased by roughly 4 percentage points from 1940 to 1990. The largest fluctuations in the wage gap occurred after 1970. During the 1970s the wage gap decreased by 6 percentage points. The decomposition as a whole during the 1970s does a poor job explaining the decrease in the wage gap during this period. The skill price of education increased the wage gap by just less than one percent during the 1970s. Changes in the prices to unobserved characteristics also work against Appalachians, increasing the wage gap by just less than one percentage point. If the residual is interpreted as payments to unobservable skill, this term is consistent with the origin of the rising prices of unobservable skill seen in Juhn, et al. (1993).

During the 1980s the wage gap increased by over 8 percentage points. The wage convergence decomposition in the 1980s shows that changes in prices alone explain over one-third of the increase in the wage gap. Specifically, changes in education prices and occupation premia each increase the wage gap by one percentage point. Unobserved prices also increase the wage gap by one percentage point. If the residual is interpreted as payments to unobservable skill, this term coupled with the education and experience prices show an overall increase in the returns to skill were unfavorable to Appalachians, likely exacerbated by their below-average skill levels. If the residual is interpreted as differences in school quality, then changes in the quality-adjusted school price alone accounts for nearly half (0.0137 + 0.0258) of the increase in the wage gap. Many researchers have noted this increase in the skill premia during this time frame (e.g., Bound & Johnson (1992), Murphy & Katz (1992), and Juhn, et al. (1993)) and have attributed this to skill biased technological change. For these results, the existence of skill biased technological change requires the assumption that at least some portion of the differences in observables and the residual are picking up differences in skill between the two samples.

The period before 1970 is less interesting because of smaller changes in the wage gap between Appalachia and the rest of the country. Wages between 1940 and 1970 grew substantially both inside of Appalachia and the rest of the country at roughly the same rate. The decrease in the wage gap during the 1940s is consistent with wage compression found in Goldin & Katz (1999) during that period. Depending on the interpretation of the residual, decreases in payments to unobserved skill or school prices (price of school plus school quality adjustment) explain nearly all of this wage gap change. There is virtually no change in the wage gap during the 1950s. The decompositions suggest changes in experience were unfavorable during this period, but the large unexplained portion of the decompositions sheds some doubt on these estimates. The wage gap decomposition of the 1960s suffers the same problem. Observable quantities and prices are in favor of Appalachians despite an increase in the wage gap of over 3 percentage points.

Comparing Appalachians to rural non-Appalachians removes differences in the urban-rural makeup between Appalachia and the rest of the country. Table 4.1 illustrates that the two groups earned roughly the same mean wage in 1940 and 1990. Rural non-Appalachians increased the wage gap each decade until 1970 when the wag gap was 9 percent. During the 1970s Appalachians virtually erased the wage gap and little change occurred during the 1980s. Taken separately, the wage means suggest that much, if not all, of the Appalachian wage gap is explained by its rural setting. Because wage gaps and means of observables between these two samples are appreciably smaller than in the Appalachian/full non-Appalachian comparison, the decompositions explain small portions of the changes wage gap in most of the decades. The exceptions are the 1940s and 1950s, when quantity changes in industry and occupation shares were largely unfavorable to Appalachians, increasing the wage gap by two percent in each decade.

4.5 Demand Shifts

One way to investigate differences between Appalachia and the rest of the country is to estimate shifts in the demand for labor in Appalachia and the rest of the country. Following Katz & Murphy (1992), I divide the regional economy of each sample into industry and occupation shares.[33] Let $E_{i,o}^{r}$ be the absolute number of worker hours in region *r*, industry *i*, and occupation *o*. Let $\Delta E_{i,o}$ be the change in the absolute number of national worker hours in industry *i*, and occupation *o,* and *E* be the total number of national worker hours. The regional change in labor demand is

$$\Delta X^{r} = \sum_{i,o} \left(\frac{E_{i,o}^{r}}{E^{r}} \right) \left(\frac{\Delta E_{i,o}}{E} \right).$$

In words, the change in the regional demand is the regional share in each industry-occupation cell multiplied by the change in national demand for each cell. The sum of this product over all industry-occupation cells estimates regional demand shifts. Because this method uses regional industry-occupation shares as weights, regions with a large employment shares in booming industries and occupations will have positive shifts in demand and the opposite will be true for regions with high shares in declining industries or occupations.

Table 4.10 presents the demand shift parameters. The three sets of results correspond to the three different aggregations of regional demand: industry, occupation, and industry and occupation combined. The Appalachian demand change is negative and the non-Appalachian demand change is positive for each decade in the sample after 1950. Further investigation into the industry and occupation shares shows the largest Appalachian industries and occupations are declining on the national level. It is likely that demand growth for high-skilled workers during the 1970s and 1980s (found in Katz & Murphy) were unfavorable to Appalachia because of the below-average skills of the workforce. In contrast, non-Appalachians experienced positive shifts in demand in every decade except the 1940s.

A large portion of the Appalachian decrease in demand is caused by the manufacturing industries and the operative occupation. The two manufacturing industries in the data, durable and non-durable, decreased in national employment shares from just under 26% in 1970 to just under 20% in 1990. The operative occupation has experienced similar declines in national participation rates from 19% to 12%. Compounding the problem for Appalachians is that the manufacturing industries and operative occupation have the largest participation rates in their respective categories and are considerably higher than the participation rates in the other two samples.

4.6 Conclusions

The Appalachian mean weekly wage in 1990 was 22% lower than the mean non-Appalachian wage. Although the wage gap between these two groups has fluctuated, the wage differential has grown by four percentage points since 1940. One cause for the divergence in wages is the gap in human capital. The mean Appalachian education level was roughly 0.8 years lower than the mean non-Appalachian education level for each sample year from 1940 to 1990. My previous work also finds Appalachian deficits in the Armed Forces Qualifying Test (AFQT) score. The gap in AFQT would be included in the wage residual since this variable is unavailable in Census data. This gap in human capital is particularly detrimental to Appalachians during the 1980s, when returns to high levels of education increased substantially. Another factor of wage growth is the industrial and occupational composition within and outside of Appalachia. The sample means show lower Appalachian participation rates in the highest paying industries and occupations than for workers elsewhere in the country. Another factor working against Appalachians is the increase in income inequality since the 1950s, particularly the 1980s. The spread of wages over this time period works against any groups with means wages on the left side of the national wage distribution.

My results show that changes in the price of both observable and unobservable skill were detrimental to Appalachians during the 1980s, perhaps driven by skill biased technological change. This result is in line with similar decompositions done for disadvantaged groups such the blacks (Juhn, et al. (1991)), females (Blau & Kahn), and Mexicans (Trejo). The effect of school quality on the wage residual corroborates this result. Smaller changes in the wage gap pre-1970 make these results more difficult to interpret. Income inequality works against Appalachians in every decade except the 1940s, when wage compression helped decrease the wage gap. Changes in occupation shares work against Appalachians in every decade. Estimated labor demand shifts show that changes in national industry and occupation makeup have largely been unfavorable to Appalachians. If the college-high school relative wage continues to rise as it did throughout the 1970s and 1980s, the Appalachian wage gap will likely continue to increase because of the Appalachian deficit in skills.

It is also possible that the wage gap between Appalachia and the rest of the country is caused by differences in urban-rural makeup between the regions. This effect is exacerbated by the sample definition of Appalachia used in this paper, which overstates the rural composition of Appalachia. Over 90% of the Appalachian sample in this paper was living in Census-defined rural area, compared to 30% of the non-Appalachian sample. A comparison of wages between Appalachians and a sample of rural non-Appalachians show strikingly similar wages in 1990. In fact, Appalachian wages in 1940 were 8% higher than rural non-Appalachians. However, differences between these two samples do exist. Education levels for rural-non-Appalachians are higher than in Appalachia. Rural non-Appalachians have higher participation rates in the two highest paying occupations throughout the sample frame. Because the means in wages and observable characteristics are closer than using the full non-Appalachian control group, the wage gap decomposition do not reveal much information. However, the wage means alone illustrate that the Appalachian region may be symptomatic of a larger urban-rural differences in wages.

	Appalachians	Non-Appalachians	Rural non-Appalachians
1940	1.608 (0.678)	1.702 (0.659)	1.521 (0.665)
1950	1.908 (0.572)	2.007 (0.540)	1.883 (0.578)
1960	2.099 (0.642)	2.270 (0.591)	2.166 (0.605)
1970	2.271 (0.630)	2.440 (0.676)	2.347 (0.690)
1980	2.245 (0.611)	2.358 (0.640)	2.230 (0.623)
1990	2.197 (0.649)	2.394 (0.696)	2.197 (0.661)

Note: Standard deviations are in parentheses.

Table 4.1 - U.S. Mean Real Log Hourly Wages Levels by Year

	1940	1950	1960	1970	1980	1990
Education						
App	9.6	9.9	10.4	11.1	11.9	12.5
Napp	10.1	10.5	11.1	11.8	12.7	13.3
RNA	10.0	10.3	10.9	11.6	12.4	12.9
Experience						
App	2.60	2.67	2.80	2.78	2.51	2.53
RNA	2.59	2.66	2.80	2.78	2.44	2.49
Napp	2.60	2.71	2.79	2.73	2.42	2.41
Black						
App	0.056	0.046	0.041	0.043	0.048	0.041
RNA	0.039	0.040	0.037	0.040	0.030	0.024
Napp	0.052	0.066	0.070	0.084	0.086	0.076
Female						
App	0.183	0.202	0.234	0.282	0.324	0.374
RNA	0.168	0.193	0.232	0.278	0.316	0.371
Napp	0.234	0.247	0.259	0.302	0.350	0.396

Note: Standard errors are in parentheses.

Table 4.2 – Means of Observable Characteristics by Year

	1940	1950	1960	1970	1980	1990
Log Weekly Wage Gap	-0.185	-0.159	-0.163	-0.196	-0.135	-0.221
No App/school interaction term						
Appalachian dummy	-0.167 (0.004)	-0.146 (0.005)	-0.141 (0.003)	-0.163 (0.003)	-0.167 (0.004)	-0.153 (0.003)
School	0.042 (0.001)	0.036 (0.001)	0.050 (0.001)	0.051 (0.001)	0.042 (0.001)	0.078 (0.001)
With App/school interaction term						
Appalachian dummy	-0.202 (0.014)	-0.164 (0.019)	-0.186 (0.011)	-0.254 (0.012)	-0.202 (0.014)	0.009 (0.018)
School*Appalachian	0.045 (0.001)	0.038 (0.002)	0.054 (0.001)	0.058 (0.001)	0.045 (0.001)	0.066 (0.001)
School*(1-Appalachian)	0.042 (0.001)	0.036 (0.001)	0.050 (0.001)	0.050 (0.001)	0.042 (0.001)	0.078 (0.001)

Table 4.3 – Selected Means from U.S. Mincerian Wage Regressions

	1940	1950	1960	1970	1980	1990
Professional & Technical	5.81	5.98	6.26	6.47	6.40	6.54
Managers, Officials & Proprietors	5.95	6.09	6.39	6.60	6.50	6.57
Clerical & Kindred	5.46	5.64	5.84	5.96	5.94	6.02
Sales Workers	5.50	5.76	6.02	6.26	6.26	6.36
Craftsmen	5.62	5.89	6.08	6.27	6.27	6.30
Operatives	5.35	5.68	5.91	6.07	6.11	6.09
Service Workers	5.04	5.43	5.57	5.82	5.85	5.88
Farmers & Farm Laborers	4.73	5.28	5.48	5.79	5.84	5.81
Other Laborers	5.21	5.59	5.82	6.00	6.04	5.99

Table 4.4 - U.S. Mean Real Hourly Wages Levels by Broad Occupation and Year

	1940			1950		
	App	NA	RNA	App	NA	RNA
Professional & Technical	5.8%	7.1%	6.7%	6.7%	8.9%	8.0%
Managers, Officials & Proprietors	9.0%	8.9%	9.3%	8.0%	8.5%	8.9%
Clerical & Kindred	11.7%	16.7%	10.8%	11.6%	16.7%	11.4%
Sales Workers	8.4%	9.5%	8.8%	7.1%	7.8%	7.7%
Craftsmen	15.9%	14.9%	14.4%	19.8%	18.9%	18.9%
Operatives	24.7%	18.9%	18.4%	26.2%	21.3%	21.1%
Service Workers	8.5%	13.4%	10.0%	7.1%	9.2%	7.9%
Farmers & Farm Laborers	6.9%	4.5%	13.5%	5.6%	3.0%	9.3%
Other Laborers	9.2%	6.1%	8.2%	7.9%	5.5%	6.9%

	1960			1970		
	App	NA	RNA	App	NA	RNA
Professional & Technical	8.8%	12.0%	11.0%	10.7%	15.6%	14.1%
Managers, Officials & Proprietors	7.1%	9.0%	8.8%	7.4%	10.0%	9.7%
Clerical & Kindred	11.9%	16.9%	12.6%	12.0%	17.7%	13.3%
Sales Workers	6.1%	7.0%	6.6%	5.0%	6.7%	5.9%
Craftsmen	20.7%	20.3%	21.1%	21.0%	18.3%	19.9%
Operatives	29.2%	20.3%	22.8%	29.3%	18.0%	20.9%
Service Workers	7.1%	8.5%	8.0%	7.6%	8.9%	8.9%
Farmers & Farm Laborers	2.9%	1.8%	4.1%	1.7%	1.3%	3.1%
Other Laborers	6.1%	4.2%	4.9%	5.2%	3.6%	4.2%

	1980			1990		
	App	NA	RNA	App	NA	RNA
Professional & Technical	12.7%	18.6%	14.8%	14.5%	20.4%	15.7%
Managers, Officials & Proprietors	9.5%	13.9%	12.2%	11.4%	14.9%	13.2%
Clerical & Kindred	14.0%	18.5%	14.0%	13.6%	17.6%	13.9%
Sales Workers	4.0%	6.1%	4.9%	3.9%	6.4%	4.5%
Craftsmen	18.9%	16.1%	18.4%	17.0%	12.5%	15.6%
Operatives	26.3%	14.1%	18.5%	23.6%	11.8%	17.7%
Service Workers	7.8%	8.4%	8.8%	8.7%	11.0%	10.1%
Farmers & Farm Laborers	1.4%	1.1%	3.9%	1.4%	1.3%	4.2%
Other Laborers	5.2%	3.2%	4.5%	5.9%	4.0%	5.0%

Table 4.5 – Participation Rates for Broad Occupation Categories, by year

	1940	1950	1960	1970	1980	1990
Agriculture, Forestry & Fishing	4.79	5.32	5.54	5.87	5.92	5.93
Mining	5.59	5.93	6.16	6.33	6.65	6.52
Construction	5.49	5.85	6.11	6.34	6.31	6.33
Manufacturing, Durable	5.54	5.83	6.14	6.29	6.31	6.38
Manufacturing, non-Durable	5.48	5.77	6.02	6.17	6.20	6.27
Transportation	5.63	5.86	6.13	6.30	6.39	6.36
Telecommunications	5.58	5.73	6.04	6.22	6.39	6.54
Utilities & Sanitary Services	5.70	5.84	6.10	6.30	6.36	6.51
Wholesale Trade	5.60	5.85	6.10	6.30	6.28	6.36
Retail Trade	5.31	5.62	5.78	6.00	5.98	5.99
Finance, Insurance, Real Estate	5.60	5.73	5.99	6.19	6.14	6.35
Business & Repair Services	5.44	5.75	6.01	6.22	6.18	6.28
Personal Services	4.94	5.34	5.46	5.73	5.77	5.81
Entertainment & Rec. Services	5.50	5.82	6.03	6.21	6.14	6.21
Professional Services	5.47	5.68	5.90	6.12	6.12	6.29
Public	5.66	5.82	5.92	6.17	6.19	6.32

Table 4.6 - U.S. Mean Real Hourly Wages Levels for Broad Industry, by year

	1940			1950		
	App	NA	RNA	App	NA	RNA
Agriculture, Forestry & Fishing	7.4%	5.0%	14.6%	5.8%	3.4%	10.2%
Mining	6.5%	1.2%	3.4%	5.0%	1.2%	3.6%
Construction	4.2%	3.3%	4.4%	5.4%	5.3%	6.7%
Manufacturing, Durable	13.9%	13.4%	10.0%	17.5%	17.4%	12.2%
Manufacturing, non-Durable	15.1%	13.6%	10.0%	14.9%	13.4%	10.7%
Transportation	8.5%	7.4%	7.2%	7.6%	7.3%	6.8%
Telecommunications	1.2%	1.4%	1.2%	1.1%	1.7%	1.3%
Utilities & Sanitary Services	2.1%	2.1%	2.1%	2.8%	2.2%	2.5%
Wholesale Trade	2.7%	4.1%	3.7%	3.1%	4.8%	4.1%
Retail Trade	16.0%	17.5%	17.7%	14.6%	15.2%	16.6%
Finance, Insurance, Real Estate	2.7%	5.4%	2.5%	2.6%	4.5%	2.5%
Business & Repair Services	2.1%	2.4%	2.7%	3.0%	2.8%	3.2%
Personal Services	4.7%	8.1%	5.3%	3.0%	4.3%	3.5%
Entertainment & Rec. Services	0.8%	1.1%	0.8%	0.7%	0.9%	0.8%
Professional Services	6.3%	6.9%	7.4%	7.1%	7.3%	8.0%
Public	5.6%	6.9%	6.8%	5.7%	8.3%	7.4%

	1960			1970		
	App	NA	RNA	App	NA	RNA
Agriculture, Forestry & Fishing	3.5%	2.3%	5.1%	2.3%	1.8%	4.0%
Mining	4.0%	1.1%	2.5%	3.8%	1.0%	2.3%
Construction	5.8%	5.6%	6.5%	6.8%	5.9%	6.7%
Manufacturing, Durable	21.7%	20.0%	18.2%	23.5%	19.2%	18.7%
Manufacturing, non-Durable	17.2%	12.7%	11.8%	15.8%	10.0%	10.3%
Transportation	5.4%	5.6%	5.1%	4.3%	4.9%	4.1%
Telecommunications	1.2%	1.6%	1.4%	1.1%	1.6%	1.3%
Utilities & Sanitary Services	2.4%	2.0%	2.3%	2.3%	1.7%	2.1%
Wholesale Trade	3.0%	4.4%	3.7%	3.1%	5.0%	3.8%
Retail Trade	12.8%	12.8%	14.0%	11.7%	13.0%	13.8%
Finance, Insurance, Real Estate	2.7%	5.1%	3.6%	3.1%	5.7%	4.0%
Business & Repair Services	1.6%	2.7%	2.0%	1.8%	3.4%	2.3%
Personal Services	2.4%	3.2%	2.8%	1.9%	2.6%	2.2%
Entertainment & Rec. Services	0.5%	0.8%	0.6%	0.6%	0.8%	0.6%
Professional Services	8.8%	10.0%	11.1%	11.9%	14.1%	14.9%
Public	7.1%	10.1%	9.4%	5.9%	9.4%	9.0%

Continued

Table 4.7 – Participation rates for broad industry categories, by year

Table 4.7 continued

	1980			1990		
	App	NA	RNA	App	NA	RNA
Agriculture, Forestry & Fishing	1.8%	1.6%	5.1%	2.2%	2.2%	6.1%
Mining	5.3%	1.3%	3.8%	3.1%	0.9%	2.4%
Construction	5.2%	5.4%	6.7%	6.1%	5.9%	6.6%
Manufacturing, Durable	22.3%	17.3%	16.1%	20.7%	13.3%	14.7%
Manufacturing, non-Durable	14.6%	8.3%	9.0%	13.3%	7.4%	8.5%
Transportation	3.9%	4.4%	4.1%	4.0%	4.1%	4.2%
Telecommunications	1.2%	1.8%	1.4%	0.8%	1.3%	0.7%
Utilities & Sanitary Services	2.5%	1.9%	2.5%	2.2%	1.7%	2.3%
Wholesale Trade	3.6%	5.2%	4.1%	3.5%	5.2%	3.8%
Retail Trade	10.9%	12.0%	12.5%	12.0%	13.0%	13.6%
Finance, Insurance, Real Estate	3.5%	6.9%	4.5%	4.0%	7.8%	4.7%
Business & Repair Services	2.0%	4.5%	2.6%	2.8%	5.5%	3.0%
Personal Services	1.0%	1.8%	1.6%	1.5%	2.2%	2.0%
Entertainment & Rec. Services	0.5%	1.0%	0.7%	0.7%	1.5%	1.1%
Professional Services	14.4%	16.8%	16.4%	16.7%	19.3%	18.3%
Public	7.3%	10.1%	9.0%	6.3%	8.8%	8.1%

	1940-1950	1950-1960	1960-1970
Δ log wage gap	-0.0257	0.0045	0.0329
All Observables	-0.0008	0.0230	-0.0142
Quantities	0.0136	0.0229	-0.0050
Education	0.0037	0.0048	0.0032
Experience	0.0023	0.0218	-0.0026
Black	-0.0083	-0.0016	-0.0052
Female	0.0009	-0.0016	0.0005
Industry	0.0092	-0.0064	-0.0073
Occupation	0.0057	0.0058	0.0065
Prices	-0.0144	0.0001	-0.0092
Education	-0.0026	0.0077	0.0002
Experience	-0.0002	-0.0202	-0.0050
Black	-0.0001	0.0035	-0.0078
Female	0.0053	0.0116	-0.0047
Industry	-0.0043	-0.0032	0.0046
Occupation	-0.0126	0.0006	0.0034
Residual	-0.0249	-0.0185	-0.0471
Unobserved Prices	-0.0271	0.0159	0.0080
School Quality	-0.0231	0.0577	0.0010

Continued

Note: Δ log wage gap = all observables + residual, all observables = quantities + prices.

Table 4.8 – Non-Appalachian / Appalachian Decomposition

Table 4.8 Continued

	1970-1980	1980-1990
Δ log wage gap	-0.0609	0.0860
All Observables	0.0044	0.0354
Quantities	0.0091	0.0037
Education	0.0090	-0.0022
Experience	-0.0046	-0.0033
Black	0.0009	0.0013
Female	-0.0002	0.0002
Industry	-0.0046	0.0009
Occupation	0.0088	0.0067
Prices	-0.0047	0.0318
Education	0.0081	0.0137
Experience	-0.0014	-0.0007
Black	0.0016	0.0023
Female	0.0013	-0.0002
Industry	-0.0035	0.0042
Occupation	-0.0110	0.0123
Residual	-0.0653	0.0506
Unobserved Prices	0.0067	0.0153
School Quality	0.0007	0.0258

Note: Δ log wage gap = all observables + residual, all observables = quantities + prices.

	1940-1950	1950-1960	1960-1970
Δ log wage gap	0.0494	0.0253	0.0166
All Observables	0.0232	0.0182	0.0133
Quantities	0.0133	0.0195	0.0026
Education	0.0002	0.0045	0.0025
Experience	-0.0002	0.0007	-0.0001
Black	-0.0039	-0.0014	-0.00001
Female	-0.0014	-0.0020	0.0002
Industry	0.0077	0.0067	-0.0045
Occupation	0.0109	0.0109	0.0045
Prices	0.0098	-0.0012	0.0107
Education	-0.0029	0.0047	-0.0001
Experience	0.0003	-0.0002	-0.00004
Black	-0.0003	0.0004	0.00002
Female	-0.0014	0.0003	-0.00009
Industry	0.0039	-0.0116	0.0098
Occupation	0.0101	0.0051	0.0010
Residual	0.0262	0.0071	0.0033
Unobserved Prices	-0.0036	0.0007	0.0037
School Quality	-0.0040	0.0018	-0.0002

Continued

Note: Δ log wage gap = all observables + residual, all observables = quantities + prices.

Table 4.9 – Rural non-Appalachian / Appalachian Decomposition

Table 4.9 continued

	1970-1980	1980-1990
Δ log wage gap	-0.0716	-0.0048
All Observables	-0.0118	-0.0010
Quantities	-0.0034	-0.0078
Education	-0.0014	-0.0061
Experience	-0.0056	0.0023
Black	0.0062	-0.00005
Female	0.0007	-0.0009
Industry	0.0007	0.0035
Occupation	-0.0041	-0.0066
Prices	-0.0083	0.0068
Education	0.0023	0.0065
Experience	-0.00001	-0.0005
Black	-0.0001	-0.0011
Female	-0.0034	0.00008
Industry	-0.0039	-0.00006
Occupation	-0.0063	0.0019
Residual	-0.0598	-0.0038
Unobserved Prices	0.0035	0.0012
School Quality	0.0060	0.0017

Note: Δ log wage gap = all observables + residual, all observables = quantities + prices

Summed Over Occupation Only

	40-50	50-60	60-70	70-80	80-90
App	2.03%	-0.98%	-2.55%	-4.10%	-3.04%
Rural Non-App	-0.22%	-0.40%	-0.68%	-0.14%	-0.80%
Non-App	2.00%	1.23%	1.15%	1.92%	1.24%

Summed Over Industry Only

	40-50	50-60	60-70	70-80	80-90
App	2.32%	0.24%	-1.39%	-1.24%	-3.15%
Rural Non-App	0.07%	0.45%	0.73%	0.15%	0.82%
Non-App	2.86%	1.84%	1.63%	1.24%	1.65%

Summed Over Industry and Occupation

	40-50	50-60	60-70	70-80	80-90
App	3.19%	-0.30%	-1.50%	-2.84%	-4.58%
Rural Non-App	1.17%	1.52%	1.03%	0.0001%	-0.012%
Non-App	4.77%	3.64%	2.76%	3.12%	2.80%

Table 4.10 – Regional Demand Shifts, by year

LIST OF REFERENCES

Appalachian Regional Commission, "ARC Federal, State and Local Partners," www.arc.gov, [2001].

Appalachian Regional Commission, "Appalachian Regional Commission Annual Report 1999," www.arc.gov [2000].

Billings, Dwight, "Culture and Poverty in Appalachia: A Theoretical Discussion and Empirical Analysis," *Social Forces,* 53(2), [1974], 315-323.

Black, Dan, Terra McKinnish and Seth Sanders, "The Economic Impact of the Coal Boom and Bust," working paper, http://www-cpr.maxwell.syr.edu/pbriefs/pb18.pdf [2001].

Blau, Francine D., and Lawrence M. Kahn, " Swimming Upstream: Trends in the Gender Wage Differential in the 1980s," *Journal of Labor Economics* 15 (1) Part 1, [January 1997] p. 1-42.

Blanchard, Olivier and Lawrence Katz, "Regional Evolutions," *Brookings Papers on Economic Activity,* [1992], p. 1-61.

Blinder, Alan S., "Wage Discrimination: Reduced Form and Structural Estimates," *Journal of Human Resources* 8 (4), [1973], p. 436-455.

Borjas, George J., Stephen G. Bronars, and Stephen J. Trejo, "Self-Selection and Internal Migration in the United States," *Journal of Urban Economics* 32 [1992], p. 159-185.

Bound, John and George Johnson, "Changes in the Structure of Wages in the 1980s: An Evaluation of Alternative Explanations," *American Economic Review* 82, [1992], p. 371-392.

Card, David, "The Effect of Unions on the Structure of Wages: A Longitudinal Analysis," *Econometrica* 64 [July 1996], p. 957-979.

Card, David and Alan B. Krueger, "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States," *Journal of Political Economy* 100 (1), [1992], p. 1-40.

Cushing, Brian and Cynthia Rogers. 1995. "Income and Poverty." *Socio-Economic Review of Appalachia*. www.rri.wvu.edu/wpapers/pdffiles/cushingarc.pdf.

Dickens, William T., and Lawrence F. Katz, "Inter-Industry Wage Differences and Theories of Wage Determination," NBER Working Paper #2271, [1987].

Feser, Edward and Harvey Goldstein, "Regional Technology Assets and Opportunities: The Geographic Clustering of High-Tech Industry, Science, and Innovation in Appalachia," Socio-Economic Review of Appalachia (papers commissioned by the Appalachian Regional Commission), www.arc.gov, [2002].

Gabriel, Paul E. and Susanne Schmitz, "Favorable Self-Selection and the Internal Migration of Young White Males in the United States," *The Journal of Human Resources* 30 (3), [1995], p. 460-471.

Greene, William H., "Sample Selection as a Specification Error: Comment," *Econometrica* 49, [1981], p. 795-798.

Goldin, Claudia, and Lawrence F. Katz, "The Returns to Skill in the United States Across the Twentieth Century," NBER Working Paper #7126.

Hanushek, Eric, "Regional Differences in the Structure of Earnings," *Review of Economics and Statistics,* 55(2), [1973], p. 204-213.

Heckman, James, "Sample Selection as a Specification Error," *Econometrica* 47(1), [1979], p. 153-161.

Hersch, Joni, "Compensating Differentials for Gender-Specific Job Injury Risks," *American Economic Review* 88 (3), [1998], p. 598-607.

Isserman, Andrew M.,. "Appalachia Then and Now: An Update on the Realities of Deprivation" Report to the President in 1964," *Journal of Appalachian Studies,* 3(1), p. 43-69 [1997].

Isserman, Andrew M., "The Evolving Appalachian Economy," http://www.arc.gov/research/pdf/evolve.pdf [1996].

Jovanovic, Boyan, "Job Matching and the Theory of Turnover," *Journal of Political Economy* 87, [October 1979], p. 972-990.

Jovanovic, Boyan, “Firm-Specific Capital and Turnover,” *Journal of Political Economy* 87, [December1979], p. 1246-1260.

Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce, “Accounting for the Slowdown in Black-White Wage Convergence.” In Workers and Their Wages, edited by Marvin Kosters, p. 107-143, Washington, D.C.: AEI Press, [1991].

Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce, “Wage Inequality and the Rise in Returns to Skill,” *Journal of Political Economy* 101 (3), [1993], p. 410-442.

Kaboski, Joseph P., “Growth Accounting and Job Quality,” working paper, Ohio State University, [2002].

Katz, Lawrence F., and Kevin M. Murphy, “Changes in Relative Wages, 1963-87: Supply and Demand Factors,” *Quarterly Journal of Economics* 107, [1992], p. 35-78.

Krueger, Alan B., and Lawrence H. Summers, “Efficiency Wages and the Inter-Industry Wage Structure,” *Econometrica* 56, [March 1988], p. 259-293.

Lee, Lung-fei, “Unionism and Relative Wage Rates: A Simultaneous Equations Models with Qualitative Limited Dependent Variables,” *International Economic Review* 19(2), [1978], p. 415-533.

Massey, Douglas S., “The Age of Extremes: Concentrated Affluence and Poverty in the Twenty-First Century,” *Demography* 33 (4), [1996], p. 395-412.

McFadden, Daniel, “A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration,” *Econometrica* 57 (5), [1989], p. 995-1026.

Murphy, Kevin M., and Robert Topel, “Unemployment, Risk and Earnings,” in K. Lang and J. Leonard (eds.), *Unemployment and the Structure of Labor Markets*, Oxford Press.

Murphy, Kevin M., and Finis Welch, “Industrial Change and the Rising Importance of Skill,” *American Economic Review* 83 (2), [1993], p.104-109.

Neal, Derek A. and William R. Johnson, “The Role of Premarket Factors in Black-White Wage Differences,” *Journal of Political Economy* 104 (5), [1996], p. 869-895.

Neumark, David, “Employers’ Discriminatory Behavior and the Estimation of Wage Discrimination,” *The Journal of Human Resources* 23 (3), [1988], p. 279-295.

Oaxaca, Ronald, "Male-Female Wage Differentials in Urban Labor Markets," *International Economic Review* 14 (3), [1973], p. 693-709.

Polachek, S.W. and F. Horvath, "A Life Cycle Approach to Migration: Analysis of the Perspicacious Peregrinator," *Research in Labor Economics*, vol. 1, ed. R. Ehrenberg, [1977], Greenwich, Connecticut: JAI Press.

Raphael, Steven and David A. Riker, "Geographic Mobility, Race, and Wage Differentials," *Journal of Urban Economics* 45(1), [1999], p. 17-46.

Rogers, Cynthia and Brian Cushing, "Income and Poverty," Socio-Economic Review of Appalachia (papers commissioned by the Appalachian Regional Commission), www.arc.gov [1996a].

Rogers, Cynthia and Brian Cushing, "The Appalachian Labor Force," Socio-Economic Review of Appalachia (papers commissioned by the Appalachian Regional Commission), www.arc.gov [1996b].

Sjaastad, L.A., "The Cost and Returns of Human Migration," Journal of Political Economy 70 (5), [1962], p 80-93.

Solon, Gary, "Intergenerational Income Mobility in the United States," American Economic Review 82(3), [1992], p. 393-409.

Stern, Stephen, "Two Dynamic Discrete Choice Estimation Problems and Simulation Method Solutions," The Review of Economics and Statistics 76 (4), [1994], p. 695-702.

Trejo, Stephen J., "Why Do Mexican Americans Earn Low Wages?" *Journal of Political Economy* 105 (6), [1997], p. 1235-1268.

Viscusi, Kip W., and Michael J. Moore, "Worker Learning and Compensating Differentials," *Industrial and Labor Relations Review*, Vol. 45, no. 1 [1991], pp. 80-96.

Weiler, Stephan, "Unemployment in Regional Labor Markets: Using Structural Theories to Understand Local Jobless Rates in West Virginia," *Industrial and Labor Relations Review* 54(3), [2001], p. 573-592.

Wood, Lawrence and Gregory Bischak, "Progress and Challenges in Reducing Economic Distress in Appalachia: An Analysis of National and Regional Trends Since 1960," www.ARC.gov/research/resmain.htm [2000].

Zimmerman, David J., "Regression Toward Mediocrity in Economic Stature," American Economic Review 82(3), [1992], p. 409-429.

ENDNOTES

[1] Source: Author's calculations from IPUMS 1940 and IPUMS 1990.

[2] Source: Appalachian Regional Commission.

[3] Source: 1990 Census. White poverty rates include non-black Hispanics, thereby understating the importance of Appalachia as a reservoir of concentrated non-black, non-Hispanic poverty.

[4] Sections from the second chapter are taken from the unpublished manuscript "Employment, Earnings, and Migration of Economically Disadvantaged Whites: Understanding Why Appalachians Are Particularly Disadvantaged" by R.W. Baumann and Patricia Reagan.

[5] Raphael and Riker (1999) and Polachek and Horvath (1977) use a 2-step procedure replacing actual migration with predicted value of migration from a probit.

[6] All dollar values are measured in constant 1990 dollars.

[7] Source: U.S. Census 1990.

[8] There is ample evidence that blacks in Appalachia are disadvantaged relatively to whites. For example, in 1990 according to the Census the black poverty rate in Appalachia was 32.4 percent compared to a regional white poverty rate of 13.5 percent.

Black poverty rates in Appalachia are about 10 percent higher than they are in the rest of the country. While the question of economic disadvantage among blacks in Appalachia is important, it is outside the scope of this paper.

[9] Source: Appalachian Regional Commission Annual Report 1999. Amount is reflective of spending through fiscal year 2003.

[10] The high variance in growth is a byproduct of the rise and fall of the coal industry in the 1970s and 1980s. Although the coal industry's effect on the region is an important topic, it is beyond the scope of this paper. Black, McKinnish & Sanders (2001) offer an interesting look on the human capital spillovers from the coal demand shocks during this period.

[11] Several works divide Appalachia into three regions: northern, central, and southern. The northern region is the most industrialized and stretches into the Rust Belt of western Pennsylvania and New York. The central region is the poorest and most mountainous subdivision. The southern region has experienced an economic rebound partially due to higher levels of urbanization experienced in the American South.

[12] If individual i has already migrated from the original labor market in period t, I fix the local labor market conditions at the values that obtained during the period he was last observed living there. In contrast, I use actual observations of post migration time varying individual characteristics.

[13] The reservation wage is the highest wage at which the individual is indifferent between working and not working. The individual will enter the labor force if offered a wage in

excess of the reservation wage and will remain outside the labor force if offered a wage below the reservation wage.

[14] More than 75 percent of the sample comes from the poor white oversample of the NLSY79. This group was not interviewed after 1990. Hence, we use only the 1979-1990 waves of the NLSY79.

[15] The AFQT test was administered in 1980, when some of the respondents were under the age of 18. The AFQT test is designed and calibrated for people age 18 and over, since it was designed for military recruits. Younger NLSY respondents scored systematically lower than those who were 18 or over at the time of the test. Thus, age adjusted AFQT scores are used. Birth year means are calculated using the cross section sample to create a benchmark. The age-adjusted scores are differences between the individual's AFQT score and the birth year benchmark.

[16] The average distance between county centroids is about 20 miles. Thus the move measure can be thought of as requiring a move at least three adjacent counties apart.

[17] These figures take the individual as the unit of observation and do not pool observations across years.

[18] All dollar values are expressed in 1990 dollars.

[19] The unemployment rates come from the Bureau of Labor Statistics Local Area Unemployment Rate series. The data on county earnings in the retail and manufacturing sectors comes from County Business Patterns.

[20] Data from the 1956 and 1962 City and County Data Book are used to calculate the percent urban in county of birth variable.

[21] An individual is considered to be at risk of migration in all years if he is never observed to migrate or in the years before migration if he is observed to migrate.

[22] It is plausible that the higher premium for Appalachians might be a return for longer distance moves because of the different definition of migration for Appalachians and non-Appalachians. However, the average distance of moves for Appalachians is 425, substantially lower than the average distance of moves for non-Appalachians.

[23] See Jovanovic (1979a,1979b).

[24] $1.45*0.08 + 13.33*+0.004 = 11.6532$

[25] Heckman (1979) and Greene (1981) both illustrate the effect on the standard errors of the two-step Heckman estimator.

[26] Author's calculation from IPUMS 1990.

[27] The IPUMS variable metropolitan area lists urban areas. In these states they are Cincinnati, Dayton, Toledo, Columbus, Akron, Cleveland (OH), New York City, Albany, Syracuse, Buffalo, Rochester, Binghamton (NY), Atlanta, Columbus, Athens, Albany, Macon, Savannah (GA), Baltimore, Washington DC metro area (Maryland portion), Jackson, Vicksburg (MS), Richmond, Washington DC metro area (Virginia portion), Virginia Beach, Roanoke (VA), Louisville, Newport, Covington, and Lexington (KY).

[28] This population minimum is time-dependent. See www.ipums.org for further detail.

[29] Binghamton, NY is the one urban area within the Appalachian region included in the non-Appalachian sample. Urban New York was considered non-Appalachian because of the several large urban areas outside of the region including New York City, Buffalo, and Albany.

[30] The IPUMS metropolitan variable makes four distinctions: living in a central city, living outside a central city, central city status unknown, (all three imply the respondent is living in a metro area), and not living in a metro area.

[31] The choice of using control group prices reflects the evaluation of discrimination in their model in the Juhn et al. (1991) model. Since discrimination is not a factor in this model, it assumes that Appalachians and non-Appalachians are facing the same skill prices. Another option is to use skill prices from pooled regressions a la Neumark (1988). However, the pooled estimates of skill prices are strikingly close to the non-Appalachian skill prices because (i) the pooled sample is predominantly non-Appalachian and (ii) the skill prices between the groups are similar. Thus, using the pooled estimates does not change the results in any substantial way.

[32] It is also possible to solve for wage convergence with the quantity effect using prices from period *t* and the prices effect using the gap in observable characteristics from period *t'*.

[33] Katz & Murphy (1992) take the additional step by estimating demand shifts for combinations of gender, education levels, industry and occupation.